Structural characterization of eukaryotic mRNA decay factors involved in post-transcriptional gene regulation

DISSERTATION

der Mathematisch-Naturwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

vorgelegt von Andreas Boland

aus Köln

Tübingen

2013

Tag der mündlichen Qualifikation: 25.6.2014

Dekan: Professor Dr. Wolfgang Rosenstiel

1. Berichterstatter: Professor Dr. Thilo Stehle

2. Berichterstatter: Professor Dr. Elisa Izaurralde

This thesis describes work carried out in the laboratory of Prof. Dr. Elisa Izaurralde at the Max Planck Institute for Developmental Biology Tübingen, Germany, from January 2009 until November 2013. The work was supervised by Prof. Dr. Thilo Stehle at the Eberhard Karls University Tübingen, Germany, and was supported by an International PhD program Tübingen fellowship. I declare that this thesis is the product of my own work. Wherever parts of the work have been published, wherever other sources have been used as well as wherever parts of the work were done by colleagues of mine, this has been indicated accordingly.

Table of contents

1 Summary

Throughout my graduate work I was highly interested in the structure and function of protein complexes involved in the post-transcriptional regulation of gene expression, in particular in those which are involved in RNA interference (RNAi) and deadenylation.

RNAi is an evolutionarily conserved mechanism that is triggered by different classes of small RNAs (siRNAs, miRNAs, piRNAs), which associate with one member of the Argonaute family of proteins to form RNA-induced silencing complexes (RISCs). RISCs utilize these small RNAs as guides to silence sequence specific mRNA targets either by inducing their degradation and/or translational repression. Degradation of mRNA targets can be achieved directly through endonucleolytic cleavage by the Argonaute protein or indirectly by facilitating the recruitment of the CCR4-NOT and PAN2-PAN3 deadenylation complexes via the interaction with an essential bridging factor termed GW182 (or TNRC6 proteins in vertebrates). Members of the GW182 protein family interact with a subset of Argonaute proteins and also bind directly to both deadenylase complexes. Deadenylated mRNAs in turn can be decapped and are then degraded by the 5′-3′ exonuclease XRN1 in the miRNA-mediated gene silencing pathway.

In the first half of my thesis, I solved the crystal structures and characterized both the MID domain in isolation as well as the MID-PIWI lobe of the eukaryotic Argonaute protein QDE-2 from the filamentous fungi *Neurospora crassa*. The high resolution structures allowed the identification and characterization of certain eukaryotic specific insertions, such as a helix-turn-helix motif which is absent in prokaryotic counterparts but its presence is highly conserved in eukaryotes. In addition, by using RNA binding studies, I was able to demonstrate that in eukaryotic Argonaute proteins, the integrity of the interface between the MID and PIWI domains is crucial for RNA-guide binding.

In the second half of my thesis, I worked on the two major cytoplasmic deadenylase complexes in eukaryotes, the CCR4-NOT complex and the PAN2-PAN3 complex. I solved the crystal structure of the folded domains of the adaptor protein PAN3, which is crucial in regulating the deadenylation activity of the PAN2-PAN3 complex. The PAN3 structure led to the identification of the binding site of the catalytic subunit PAN2 and unveiled the presence of a tryptophan-binding pocket, which is important for the interaction with GW182/TNRC6 proteins.

In collaborative projects, I also determined the crystal structures of several sub-complexes of the CCR4-NOT complex, thereby uncovering molecular insights into the assembly of this large multi-subunit complex. Surprisingly, the structure of the NOT-module core (consisting of the C-terminal regions of CNOT1, CNOT2 and CNOT3) revealed a unique structural arrangement where intrinsically disordered regions (IDR) orchestrate complex assembly. The crystal structure is of particular interest because many RNA-binding proteins recruit the CCR4-NOT complex to target mRNAs, by binding to the NOT-module subunits. Furthermore, all the complex components were also crystallized in isolation, enabling us to describe structural rearrangements upon complex formation.

Finally, I solved the crystal structures of the CNOT1 MIF4G domain and the CNOT1 DUF3819 domain, in complex with DDX6 and the CNOT9 subunit of the CCR4-NOT complex, respectively. These two complexes provide insights into the recruitment of the CCR4-NOT complex to miRNA targets by GW182/TNRC6s proteins and further connect translational repression and possibly also decapping mediated by DDX6 to the CCR4-NOT complex.

In essence, my work provides an extensive structural and functional framework on the assembly of the two major cytoplasmic deadenylases. Furthermore, I provide structural evidence on how the deadenylases are recruited to miRNA targets. Finally, I

determined the first high-resolution crystal structure of the MID-PIWI lobe of a eukaryotic Argonaute protein (the key factor of all RNA induced silencing pathways).

1 Zusammenfassung

Während meiner Doktorarbeit interessierte ich mich für die Struktur und Funktion von Proteinkomplexen die an der post-transkriptionellen Regulation der Genexpression beteiligt sind, insbesondere solche, die an den Prozessen der RNA Interferenz (RNAi) und der Deadenylierung beteiligt sind.

RNAi ist ein evolutionär konservertier Mechansimus, der durch verschiedene Klassen von kleinen RNAs (siRNAs, miRNAs, piRNAs) ausgelöst wird, die an ein Mitglied der Familie der Argonaute Proteinen binden, um so genannte „RNA-induced silencing complexes“ (RISCs) zu formen. RISCs benutzen diese kleinen RNAs als Wegweiser, um Sequenz-spezifische mRNAs still zu legen, entweder durch die Einleitung ihres Abbaus und/oder durch die Unterdrückung ihrer Translation. Der Abbau von Ziel-mRNAs kann entweder durch das endonukleolytische Spalten der Ziel-mRNA durch das Argonaut Protein direkt erfolgen oder indirekt durch das Rekrutieren der CCR4-NOT und PAN2-PAN3 Deadenylierungskomplexe, durch die Interaktion mit einem unverzichtbaren Brückenfaktor der GW182 (oder TNRC6 Proteine in Wirbeltieren) genannt wird. Mitglieder der GW182 Proteinfamilie interagieren mit bestimmen Mitgliedern der Argonaut Proteinfamilie und können außerdem auch direkt an die beiden Deadenylierungskomplexe binden. Deadenylierte mRNAs wiederum werden dann Entkappt und unwiederbringlich abgebaut durch die 5′-3′ Exonuklease XRN1 im miRNA-vermittelten Gen-Stillegung Stoffwechselweg.

Während der ersten Hälfte meiner Doktorarbeit löste ich die Kristallstrukturen und charakterisierte sowohl die MID-Domäne alleine, als auch den ersten eukaryotischen MID-PIWI Flügel des Argonaut Proteins QDE-2 vom fadenförmigen Pilz *Neurospora crassa*. Die hoch aufgelösten Strukturen ermöglichten die Identifizierung und Charakterisierung bestimmter eukaryotischer Insertionen, wie zum Beispiel ein Helix-Turn-Helix Motiv, welches in den prokaryotischen Pendants fehlt, aber das Vorhandensein stark konserviert ist in Eukaryoten. Zusätzlich konnte ich mit Hilfe von

RNA Bindungsstudien zeigen, dass in eukaryotischen Argonaut Proteinen die Unversehrtheit der Berührungsfläche zwischen der MID und der PIWI Domänen für die RNA Bindung notwendig ist.

Während der zweiten Hälfte meiner Doktorarbeit arbeitete Ich an den beiden wichtigsten zytoplasmatischen Deadenylierungskomplexen in Eukaryoten, dem CCR4-NOT Komplex und dem PAN2-PAN3 Komplex. Hier war es mir möglich die Kristallstruktur der gefalteten Domänen des Adapterproteins PAN3, welches entscheidend für die Regulation der Deadenylierungsaktivität des PAN2-PAN3 Komplexes ist, zu lösen. Die PAN3 Struktur führte zur Identifizierung der Bindungsstelle zur katalytischen Untereinheit PAN2 und enthüllte auch die Existenz einer Tryptophan-Bindungstasche, die wichtig ist für die Interaktion mit GW182/TNRC6 Proteinen.

In gemeinschaftlichen Projekten löste ich auch die Kristallstrukturen von verschiedenen Unterkomplexen des CCR4-NOT Komplexes und konnte dadurch molekulare Erkenntnisse über die Assemblierung dieses großen und aus vielen Einheiten bestehenden Komplexes gewinnen. Überaschenderweise enthüllte die Struktur des NOT-Moduls (welches minimal aus den C-terminalen Regionen von CNOT1, CNOT2 und CNOT3 besteht) ein einzigartigen strukturellen Aufbau, welcher zeigt, dass sogenannte intrinsisch ungeordnete Regionen (IDRs) notwendig sind für die Assemblierung des Komplexes. Diese Kristallstruktur ist von besonderem Interesse, da viele RNA-bindende Proteine den CCR4-NOT Komplex zu Ziel-mRNAs rekrutieren, indem sie an die NOT-Modul Einheiten binden. Außerdem wurden alle Komplexbestandteile auch isoliert kristallisiert, was es uns ermöglichte, strukturelle Neuanordnungen nach Komplexbildung zu analysieren.

Zuletzt löste ich die Komplexstruktur der CNOT1 MIF4G Domäne und der CNOT1 DUF3819 Domäne, in Komplex mit den zugehörigen Bindungspartnern DDX6 und der integralen Untereinheit CAF40 des CCR4-NOT Komplexes. Diese zwei

Strukturen ermöglichten Einblicke in die Rekrutierung des CCR4-NOT Komplexes zu miRNA Zielen durch GW182/TNRC6 Proteine und verbinden außerdem durch DDX6 die Unterdrückung der Translation und möglicherweise auch Entkappung der Ziel-mRNA mit dem CCR4-NOT Komplex.

Im Wesentlichen stellt meine Arbeit einen umfangreichen strukturellen und funktionellen Rahmen über die Struktur und den Aufbau der beiden zentralen zytoplasmatischen Deadenylierungskomplexe dar. Weiterhin konnte ich strukturelle Informationen gewinnen, wie die Deadenylierungskomplexe zu miRNA Zielgenen rekrutiert werden. Schließlich bestimmte ich die erste hoch-aufgelöste Kristallstruktur des MID-PIWI Flügels eines eukaryotischen Argonaut Protein (der zentrale Faktor aller RNA induzierten Stilllegungsstoffwechselwege).

2 Introduction

The concentration and the quality of mRNAs available for translation in the eukaryotic cell depends on both the rate of synthesis and the rate of decay. In recent years, the mechanisms that control mRNA decay rates have come more and more into the scientific focus, in particular since the discovery of small RNA-induced silencing pathways in the early 1990s; these pathways, together with additional post-transcriptional control mechanisms, are now known to promote rapid mRNA degradation, thereby allowing the cell to rapidly adapt to environmental changes or stress conditions. Proteins that are required only for a short period, e.g. cytokines, often have mRNAs with short half-lives (less than 1 hour), whereas house-keeping genes, in contrast, often have very stable transcripts (more than 1 day) (see Figure 1).

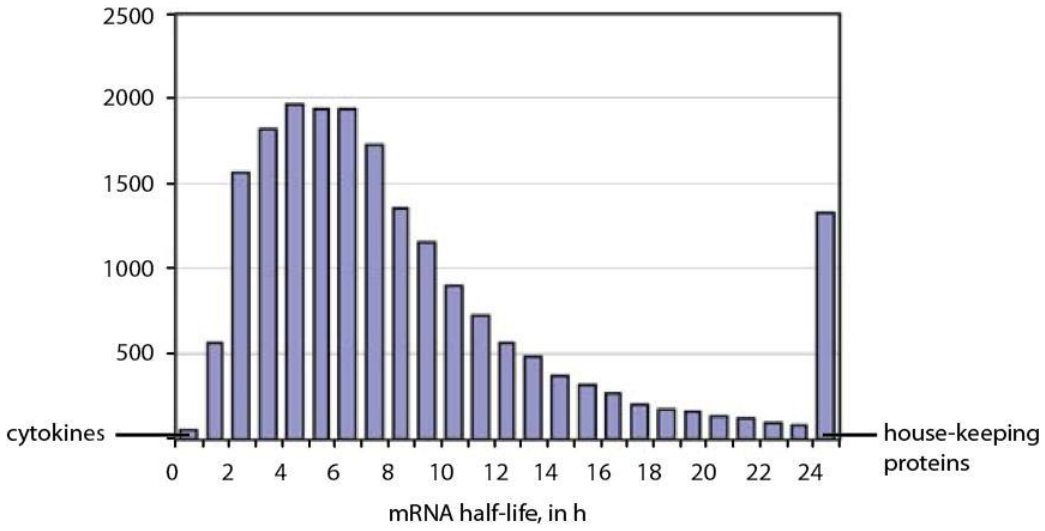


Figure 1: Frequency distribution of mRNA half-lives among all genes in mouse embryonic stem cells. (modified from Sharova et al., 2009).

Additionally, global analyses have shown that ~20-50% of changes in gene expression rather depend on variations in mRNA stability (Neff, 2013) instead of on a modification in transcription rates, further underlining the importance of these

processes. In the following chapters, I will give an overview on post-transcriptional gene regulation and the molecular mechanisms and protein complexes they depend on.

2.1 mRNA processing and translation

In the eukaryotic cell, almost all protein-coding mRNAs are modified in the nucleus before being exported into the cytoplasm for translation. These modifications include the addition of a 5′-m^7GpppN cap structure and adding a poly(A) tail at the 3′-end, a mechanism commonly known as polyadenylation; only a few mRNAs are not polyadenylated, i.e. histone mRNAs (Marzluff et al., 2008). The length of the poly(A) tail is species-dependent with an average length of approximately 70 residues in yeast and around 200-250 nucleotides in human (Eckmann et al., 2011; Wahle and Winkler, 2013). The poly(A) tail is bound with high affinity by members of a specific protein family termed poly(A) binding proteins or PABPs. This highly conserved protein family has been subdivided into two classes, nuclear and cytoplasmic, based on their intracellular localization (PABPN1 and PABPC1 are the most common representatives, respectively). The cap structure and the poly(A) tail have been shown to be important for both, the regulation of the mRNA stability (Bernstein et al., 1989; Bernstein and Ross, 1989; Filipowicz, 1978; Furuichi et al., 1977) by protecting the mRNA body from degradation through cytoplasmic exonucleases, as well as for efficient translation (see below). Moreover, non-coding regions or so-called introns are spliced out from the pre-mRNA, a process which occurs co-transcriptionally, before the mature mRNA gets exported.

Once the mRNA is exported into the cytoplasm, PABPN1 (Wahle, 1991) is replaced by its major cytosolic version PABPC1. Despite the common ability to bind poly(A) sequences, PABPN1 possesses only low structural and functional similarity to its cytosolic counterpart. Whereas PABPN1 is important for polyadenylation of the mRNA, PABC1 promotes mRNA stability and translation efficiency. Although a poly(A)

tail is not essential for translation, mRNAs lacking a poly(A) tail are less efficiently translated compared to their polyadenylated counterparts under ribosome-limiting conditions (Gallie, 1991; Munroe and Jacobson, 1990; Proweller and Butler, 1997). It has been shown that the poly(A) tail and the 5′-cap structure function synergistically to promote translation initiation (Gallie, 1991). The cap structure is bound by the eukaryotic translation initiation factor 4E (eIF4E), which also interacts with eIF4G. eIF4G in turn interacts with PABPC, leading to a circularized structure of the mRNA, the so-called closed loop conformation (see Figure 2), in which both ends of the mRNA end up in close proximity (Le et al., 1997). The simultaneous interaction of eIF4G with eIF4E and PABPC1 likely explains the synergetic effect between the cap and the poly(A) tail in translation initiation.

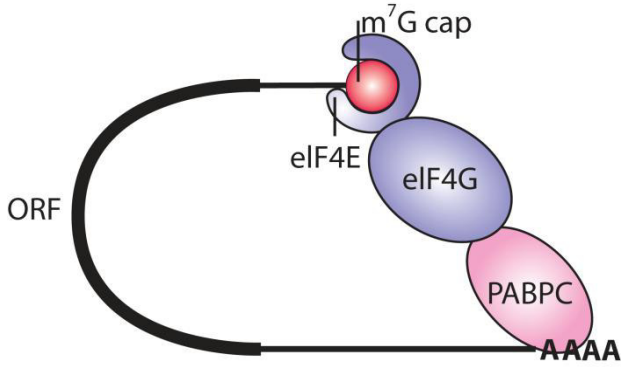


Figure 2: Closed-loop conformation of the mRNA. eIF4G bridges the interaction between the PABPC bound to the poly(A) tail of the mRNA and the eIF4E protein bound to the mRNA cap structure (adapted from Braun et al., 2013).

Because the 5′-cap structure and the poly(A) tail are stabilizing elements, an mRNA that is selected for degradation first needs to be deadenylated and decapped. These processes are described in the next two chapters.

2.2 mRNA degradation in eukaryotes

The decay of the mRNA body can be either initiated from the ends of the target RNA or through endonucleolytic cleavage as in specialized mRNA decay pathways.

However, eukaryotic bulk mRNA degradation normally begins with the shortening or removal of the poly(A) tail, a process known as deadenylation. So far, three major independent deadenylase enzymes have been identified: the CCR4-NOT complex, the PAN2-PAN3 complex (both complexes are described in detail below) and the poly(A) specific ribonuclease, referred as PARN. In addition, numerous other deadenylases are known that exist in the mammalian genome (Astrom et al., 1991; Astrom et al., 1992; Korner and Wahle, 1997). Upon deadenylation, the mRNA target can either be degraded in the 3′-5′ direction through a multi-protein complex referred as the exosome complex, or the mRNA is first decapped and then degraded in the 5′-3′ direction. In the latter case, the mRNA is decapped by the decapping protein 2 (DCP2), and decapping is assisted by several co-activators, like DCP1, the enhancer of decapping 3 (EDC3), the metazoan-specific enhancer of decapping 4 (EDC4), the DEAD box helicase 6 (DDX6, also known as RCK, Ge-1 or Me31B), Pat1 and the Lsm1-7 complex. DDX6 and Pat1 also act as translational repressors in addition to their roles as decapping activators (see below) (Chu and Rana, 2006). Once the mRNA is decapped, the uncapped 5′-end of the mRNA target exposes a 5′-phosphate, which is recognized as a substrate for the cytoplasmic 5′-3′ exonuclease XRN1. Moreover, it has been shown recently that XRN1 is actively recruited by the decapping complex to mRNA targets (Braun et al., 2012), thus adding another layer of regulation to the decapping machinery.

Several factors involved in specialized decay pathways have been identified that promote or catalyze endonucleolytic cleavage. One well-established example is the evolutionarily conserved nonsense-mediated decay (NMD) pathway, a quality control mechanism which degrades aberrant mRNAs that possess a premature translation

termination codon or PTC (Schweingruber et al., 2013). In *Drosophila melanogaster* the first step of NMD decay is endonucleolytic cleavage close to the site of the PTC, resulting in short-lived decay intermediates that are rapidly digested by XRN1 and the exosome (Gatfield and Izaurralde, 2004). NMD can also be initiated by endonucleolytic cleavage in human cells (Eberle et al., 2009; Huntzinger et al., 2008). The conserved core of this multi-subunit machinery is formed by the three up-frameshift (UPF) proteins UPF1, UPF2 and UPF3 (Conti and Izaurralde, 2005). In multicellular organisms, additional factors such as the SMG proteins (SMG1, SMG5, SMG6 and SMG7) are required for NMD. The SMG proteins (suppressor with morphogenetic effect on genitalia) were first discovered in *Caenorhabditis elegans* (Cali et al., 1999; Hodgkin et al., 1989). Out of them, only SMG6 has been shown to be an endonuclease capable to catalyze cleavage of the mRNA target (Eberle et al., 2009; Huntzinger et al., 2008).

Another noteworthy pathway that depends on endonucleolytic cleavage is the siRNA (small-interfering RNA) pathway. In this pathway, the siRNA duplex is loaded into an mRNP particle referred as RNA-induced silencing complex (RISC). The small RNA duplex is unwound and the so-called guide strand remains associated with the RISC complex. As a result, the single-stranded RNA hybridizes with its mRNA target through perfect base complementarity and endonucleolytic cleavage of the mRNA occurs. This cleavage step is performed by an RNAse H-like domain, termed PIWI domain present in Argonaute protein (see chapter 2.4.1). Post-transcriptional gene silencing (PTGS) mechanisms mediated by small RNAs are explained in detail in chapter 2.4.

Despite this variety of different mRNA decay pathways, the first and rate-limiting step in bulk mRNA degradation is the shortening of the poly(A) tail, and deadenylation is the most efficient step in controlling mRNA decay (Decker and Parker, 1993).

2.2.1 mRNA deadenylation

Nearly all general cytoplasmic mRNA decay pathways in eukaryotes are initiated by the consecutive action of two deadenylase complexes, namely the PAN2-PAN3 [poly(A) specific ribonuclease] complex and the CCR4-NOT (Carbon catabolite repressor protein 4 – Negative regulator of transcription) complex (Boeck et al., 1996; Chen and Shyu, 2011; Miller and Reese, 2012; Tucker et al., 2001; Wahle and Winkler, 2013). These deadenylase complexes are highly conserved throughout evolution; however, no homologue of the PAN2-PAN3 complex has yet been identified in plants.

Polyadenylated mRNAs are degraded in a 3′-5′ direction via hydrolysis of the C-terminal adenylate residue and as a consequence, 5′-AMP is released. Monitoring mRNA deadenylation and decay kinetics using a Tet-off promoter-driven transcriptional pulse-chase approach revealed that mammalian deadenylation consists of a biphasic distribution (Chen and Shyu, 2011; Yamashita et al., 2005). In a first phase of deadenylation, mRNAs are degraded in a distributive manner by the PAN2-PAN3 complex leading to a highly homologues length of their poly(A) tails with approximately 110 nucleotides (depending on the species). Importantly, in this first phase no degradation of the mRNA body can be observed. In addition, this phase is not essential for deadenylation, as mRNA deadenylation can be carried out solely by the CCR4-NOT complex in PAN2-PAN3-depleted cells (Tucker et al., 2001; Yamashita et al., 2005). In contrast, the second phase of deadenylation is characterized by the processive action of the CCR4-NOT complex. The lengths of the poly(A) tail of the mRNAs becomes highly heterogeneous (between 20 and 100 nucleotides) in this phase, and decay of the mRNA body occurs (Yamashita et al., 2005). The composition and function of the different subunits of the PAN2-PAN3 complex and the CCR4-NOT complex are described in detail in the following to sub-chapters.

2.2.2 The PAN2-PAN3 complex

PAN2 is the catalytic subunit of the PAN2-PAN3 complex that was originally discovered in yeast in 1992, and identified in 2004 in mammals (Lowell et al., 1992; Sachs and Deardorff, 1992; Uchida et al., 2004). PAN2 belongs to the DEDD (Asp-Glu-Asp-Asp) class of 3′-5′ exonucleases (Moser et al., 1997). The acidic residues coordinate divalent metal ions that are essential for the catalytic activity of the protein. PAN2 consists of an N-terminal WD40 domain, a ubiquitin-specific protease domain that is predicted to be inactive in its middle region (Quesada et al., 2004), and the aforementioned C-terminal exonuclease domain [see Figure 3 and (Moser et al., 1997)]. The activity of PAN2 strictly depends on its interaction with the adaptor protein PAN3.

PAN3 binds directly to PABPC through an N-terminal PAM2 (PABPC-interacting motif 2) binding motif (see Figure 3), as a result of which PAN2 is recruited to mRNA targets. Consequently, PAN2 is inactive in the absence of its co-activator. PAN3 consists of a pseudokinase domain as well as a highly conserved C-terminal domain of unknown tertiary structure (see Figure 3). Importantly, the C-terminal domain has been shown to be critical for PAN2 binding (Mangus et al., 2004). Although it has been reported that PAN3 possesses the ability to self-interact, the stoichiometry of the complex has not yet been elucidated (Mangus et al., 2004).

Recent reports unveiled a direct link between deadenylation and the miRNA-mediated gene silencing pathways through direct interactions between the GW182 (or TNRC6 proteins in vertebrates) protein family members (see below) and the deadenylase complex subunits PAN3 and NOT1 (Braun et al., 2011; Chekulaeva et al., 2011; Fabian et al., 2011). In the case of PAN3, both folded domains (pseudokinase and C-terminal domain) are necessary for the interaction with GW182, whereas the N-terminal unstructured part of the protein is dispensable (Braun et al., 2011).

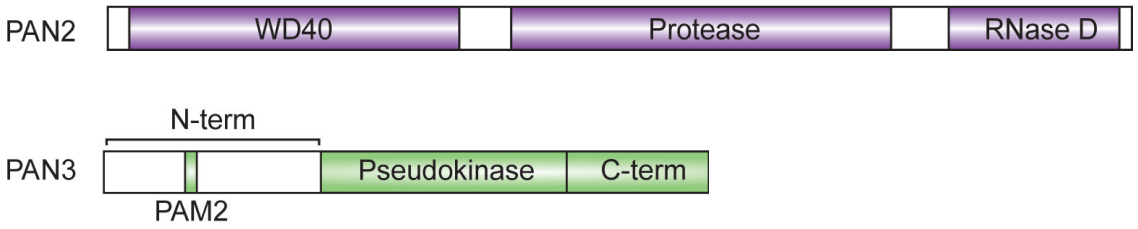


Figure 3: Domain organization of the PAN2 and PAN3 proteins. Globular domains in PAN2 and PAN3 are highlighted in purple and green, respectively. The unstructured N-terminal PAM-2 (PABP-interacting motif 2) motif present in PAN3 is also shown in green (adapted from Christie et al., 2013).

2.2.3 The CCR4-NOT complex

The CCR4-NOT complex is the major deadenylase complex in eukaryotic organisms and was originally described in 1993 as a negative regulator of transcription (Cade and Errede, 1994; Collart and Struhl, 1993; Collart and Struhl, 1994). Over the years, many functions have been attributed to the CCR4-NOT complex, including transcriptional regulation, RNA export, mRNA decay and quality control, translational repression and protein degradation induced by ubiquitylation (Collart and Panasenko, 2012; Miller and Reese, 2012; Wiederhold and Passmore, 2010). Although it is fascinating that so many functions are covered solely by this multi-subunit complex, in this thesis I will focus on the role of the CCR4-NOT complex in mRNA deadenylation.

The complex is formed out of at least five conserved core subunits; these can be divided into two modules, a catalytic module and the NOT-module. The catalytic module contains one DEDD type exonuclease, either CAF1 or POP2 and one EEP (endonuclease-exonuclease-phosphatase homology domain) type deadenylase subunit, either CCR4a or CCR4b (Bartlam and Yamamoto, 2010; Doidge et al., 2012; Wahle and Winkler, 2013). The NOT-module consists of at least NOT1, NOT2 and NOT3 or NOT5, a yeast paralogue of NOT3. Within the complex, the large subunit NOT1 serves as a modular scaffold. Additional described subunits of the complex are the ubiquitin ligase NOT4, the armadillo-repeat containing protein NOT9 (or CAF40 or RCD-1), and the

species-specific subunits NOT10, NOT11 and CAF130. Interestingly, NOT4 is part of the core CCR4-NOT complex in yeast but not in the human counterpart (Lau et al., 2009). NOT4 possesses a conserved RRM (RNA-recognition motif) domain in its C-terminus which provides potential RNA-binding capacity. Nucleic acid binding capacity has also been attributed to CNOT9, a highly conserved protein which is supposed to be involved in the regulation of cell division (Garces et al., 2007).

As mentioned above, the scaffold protein NOT1 features multiple binding sites for different sub-complexes along the entire protein. NOT1 is predicted to be almost completely α-helical with the exception of several linker regions between the domains. Except for a domain of unknown function (DUF) 3819, all domains have either been shown or are predicted to consist of primarily short α-helices that pack into HEAT-like repeats (Basquin et al., 2012; Petit et al., 2012). Recent studies have demonstrated that the binding site for the NOT10-NOT11 sub-complex resides in the N-terminus of NOT1, the CAF1-CCR4 catalytic subunits as well as the NOT9 subunit bind to the middle region, as do the NOT2 and NOT3/5 proteins to the C-terminus (Bawankar et al., 2013; Mauxion et al., 2013). The assembly of the sub-complexes is summarized in Figure 4. The CNOT1 scaffold protein has also been shown to bind to the translational repressor and decapping activator DDX6 in yeast (Maillet and Collart, 2002).

A detailed structural model for the molecular basis between the interaction of the MIF4G domain (a special case of HEAT repeats) in the middle region of NOT1 and the catalytic module of the yeast and human CCR4-NOT complex has been obtained by two recently published crystal structures (Basquin et al., 2012; Petit et al., 2012).

Furthermore, many RNA-binding proteins (RBPs) interact with the CCR4-NOT complex and here in particular with the NOT proteins, thereby recruiting the deadenylase complex to specific mRNA targets. For example, the RNA binding protein tristetraprolin (TTP) has been proven to bind to the 3′-UTR of mRNAs and to the NOT1 scaffold (Lykke-Andersen and Wagner, 2005; Sandler et al., 2011). The latter interaction

was lately determined at an atomic resolution by a co-crystal structure between a HEAT-repeat like domain of NOT1 in its N-terminus bound to a tristetraprolin peptide (Fabian et al., 2013). This mechanism, which is known as ARE-mediated decay, is described in the following chapter.

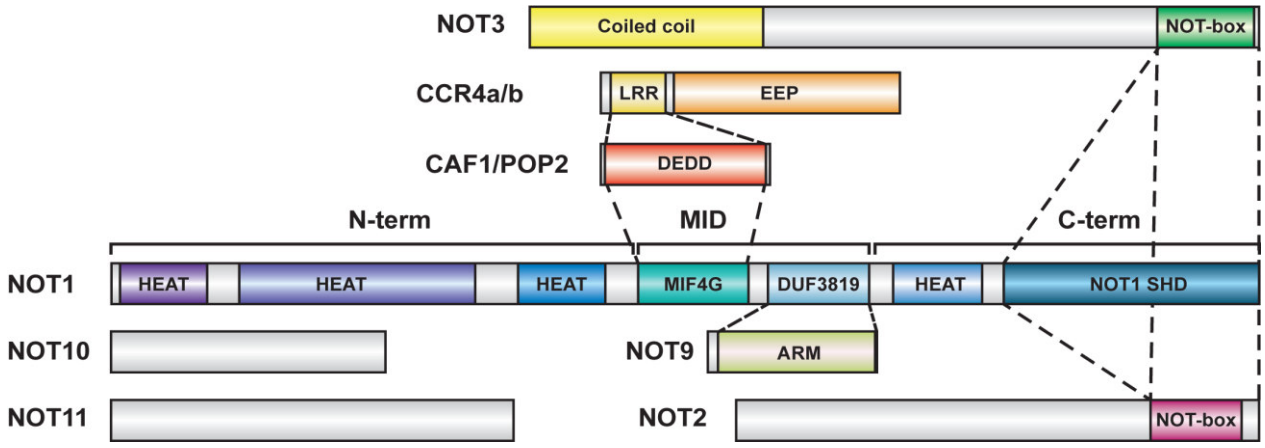


Figure 4: Domain organization of different subunits of the CCR4-NOT complex. Folded domains are colored; predicted unstructured regions are shown in gray. Dashed lines indicate known interaction between certain domains (adapted from Bawankar et al. 2013).

2.3 Destabilizing sequence elements and their binding partners

The half-life of an mRNA can vary from several minutes to more than 24 hours in mammals [(Sharova et al., 2009) and Figure 1] and is determined not only by the protecting structures such as the cap structure and the length of the poly(A) tail but, rather, by an complex interplay between those stabilizing structures and the presence of destabilizing sequence elements predominantly in the 3′-UTR of an mRNA. Tristetraprolin is a very well-known example that binds to certain destabilizing sequence elements, so-called AU-rich elements (AREs). AUUUA is a typical sequence motif that defines an ARE, which resides within a Uridine-rich sequence context in the 3′-UTR of an mRNA. These motifs are found in approximately 9% of all cellular mRNAs (Chen et al., 2001) and serve as binding platforms for proteins such as TTP which in turn

can then interact with the deadenylase machinery and thereby induce rapid mRNA decay. Thus, AREs are typically found in mRNAs that require fast mRNA turnover, such as the tumor necrosis factor α mRNA (Sandler and Stoecklin, 2008).

Another remarkable example of destabilizing elements present mainly in the 3′-UTR of target mRNAs are miRNA binding sites. These binding sites are used by small non-coding RNAs such as miRNAs (but also siRNAs and piRNAs; see below) to guide the miRISC to the 3′-UTR of target mRNAs. The miRISC complex can then induce translational repression and/or mRNA degradation. More than 1500 miRNA precursors have been identified being encoded in the mammalian genome, although the biological relevance of many of those potential miRNAs needs yet to be determined (Ameres and Zamore, 2013). It has been proposed that miRNAs regulate the gene expression for up to 30% of all human genes (Filipowicz et al., 2008).

2.4 Small RNA-mediated gene silencing mechanisms

RNA interference or RNAi was first described in the early 1990s by Jorgensen and co-workers in petunia when trying to deepen the flowers′ purple (Napoli et al., 1990). This work demonstrated that the transgenic introduction of the chalcone synthase gene, which is the key enzyme for flavonoid, flavone and anthocyanin synthesis, unexpectedly resulted in a block of anthocyanin biosynthesis, thereby leading to a partial or total loss of the flowers′ color in a wild-type pigmented background. Approximately one decade after this initial discovery, work by several groups (Elbashir et al., 2001; Fire et al., 1998; Hamilton and Baulcombe, 1999) identified a number of related and highly conserved pathways from plants to animals that completely revolutionized our understanding of gene expression and gene regulation in metazoans.

Central to all gene silencing pathways are small non-coding RNAs, roughly ~20-30 nucleotides long (miRNA, siRNA, piRNA), that interact with a member of the

Argonaute protein family to direct the resulting ribonucleoprotein complex (RNA-induced silencing complex or RISC) to specific gene targets through perfect or partial base complementarity. The small double-stranded RNA associates with the Argonaute (AGO) first, whereupon the duplex is unwound and the so-called passenger strand is then removed. In contrast, the guide strand remains associated with the protein (Guo and Lu, 2010). In the case of perfect complementarity between the guide RNA and the target strand, endonucleolytic cleavage of the target RNA can occur if the Argonaute protein harbors catalytic activity. This process is known as slicing and only a subset of Argonautes possesses all the catalytic residues and loop elements that are required for this process. Slicing is the predominant process of exogenous and endogenous small-interfering RNAs (siRNAs) and plant microRNAs (miRNAs).

In sharp contrast, animal miRNAs are usually only complementary to their targets within nucleotides 2-7 of the guide RNA, a region known as seed sequence. Downstream of the seed sequence, animal miRNAs often exhibit mismatches and bulges which prevent the cleavage of target RNAs (Liu et al., 2004; Tolia and Joshua-Tor, 2007). In this case, even catalytically active AGOs cannot cleave the target but instead facilitate the recruitment of the CCR4-NOT and PAN2-PAN3 deadenylation complexes through the interaction with a bridging factor termed GW182 (or TNRC6 proteins in human; see Figure 5). This interaction has been shown to be direct and is mediated via conserved tryptophan-rich motifs present in the silencing domain (see Figure 7). Thus GW182 proteins provide a link between miRNA target recognition by the RISC complex and subsequent target degradation by the major cytoplasmic deadenylase complexes (Braun et al., 2011; Chekulaeva et al., 2011; Fabian et al., 2011). Furthermore GW182 proteins have been shown to trigger also translational repression that precedes target degradation, but this mechanism is not yet well understood (Bazzini et al., 2012; Bethune et al., 2012; Djuranovic et al., 2012; Zekri et al., 2009). Argonaute proteins as

well as GW182 protein family members are described in more detail in the following two sub-chapters.

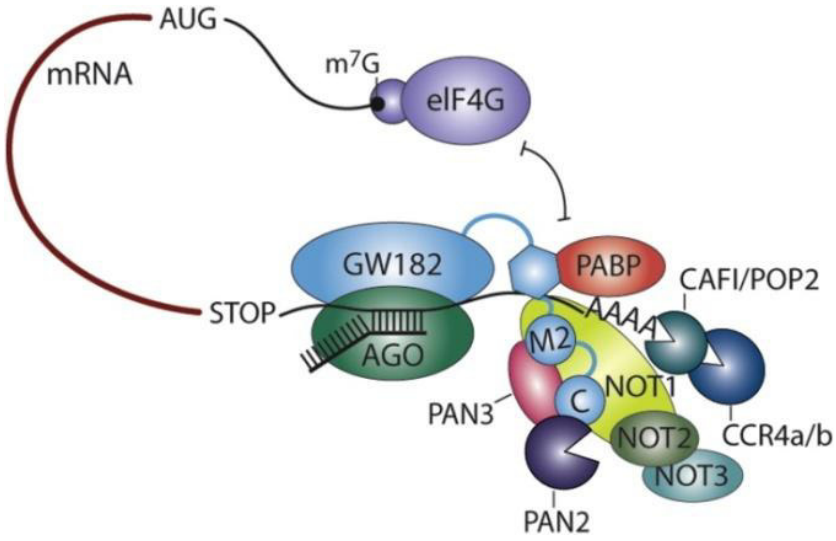


Figure 5: GW182/TNRC6 proteins recruit the two major deadenylases to miRNA targets. Recruitment is achieved by interactions with the NOT1 subunit of the CCR4-NOT deadenylase complex and with the PAN3 subunit of the PAN2-PAN3 deadenylase complex (copied from Braun et al. 2013).

2.4.1 The Argonaute proteins

Argonaute proteins (AGOs) were first described in 1998 being important for the regulation of plant development in *Arabidopsis thaliana* (Bohmert et al., 1998; Moussian et al., 1998). The unusual morphological appearance of plant AGO1 mutants reminded the authors of a small squid, which prompted them to term these mutants after the famous warriors from Greek mythology that were sailing on the ship Argo (Bohmert et al., 1998).

Argonaute proteins have been conserved from bacteria to humans throughout evolution. Although only a few studies have been published regarding the function of Argonaute proteins in bacteria and archaea (Makarova et al., 2009; Olovnikov et al., 2013; Shabalina and Koonin, 2008), their eukaryotic counterparts have diverse but very well-studied functions. Eukaryotic AGOs can act either in the cytosol together with

siRNAs and miRNAs to regulate post-transcriptional gene regulation or in the germ line together with piRNAs (PIWI-interacting RNAs) to control retrotransposons and other mobile genetic elements (Grivna et al., 2006; Lau et al., 2006). AGOs can also act in the nucleus and induce heterochromatin formation, to regulate gene expression at the level of transcriptional repression (He et al., 2009; Irvine et al., 2006). Moreover the number of AGOs differs strongly between different species. While the number of AGOs is low in prokaryotes, about 10 AGOs are found in *Arabidopsis thaliana*, more than two dozen AGOs are present in *C. elegans* genomes and only 4 cytosolic AGOs exist in humans (Pratt and MacRae, 2009; Tolia and Joshua-Tor, 2007).

Despite the fact that bacteria and archaea have evolved a different nucleic acid-based immune system (termed CRISPR system) that does not rely on Argonaute proteins, the first full-length structure was determined from the hyperthermophilic archaea *Pyrococcus furiosus* (Song et al., 2004). In the years since then, a number of full-length structures in complex with guide and target RNAs has been solved from various species namely *Archaeoglobus fulgidus*, *Aquifex aeolicus* and *Thermus thermophilus* (Ma et al., 2005; Parker et al., 2005; Wang et al., 2008a; Wang et al., 2008b; Yuan et al., 2005). These structures revealed a four domain protein consisting of a bilobal architecture with the N-terminal lobe containing the N-terminal (or simply N) and the PAZ domains, and a C-terminal lobe composed of the MID and the PIWI domains (see Figure 6).

The N-terminal domain is the least conserved part of the protein and has been shown to play an important role in the unwinding step of the RNA duplex (Kwak and Tomari, 2012). Furthermore the N-terminal domain has been identified to be an important binding platform for RISC-interacting proteins (Phetrungnapha et al., 2013). The region between the N-terminal domain and the PAZ domain has been proven to act as a hinge, and it was suggested that this motion could be beneficial for RISC loading and mRNA target binding (Ming et al., 2007; Song et al., 2004).

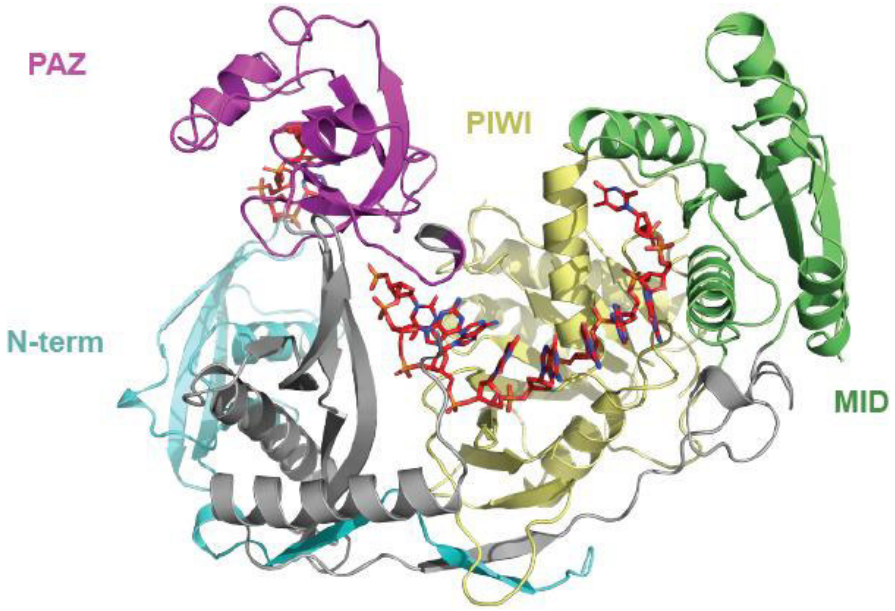


Figure 6: Crystal structure of the *Thermus thermophilus* Argonaute in complex with a 5′-phosphorylated guide DNA strand (PDB code 3DLH). The N-terminal domain and the PAZ domain are colored in cyan and purple respectively. The MID and the PIWI domain are shown in green and in yellow.

The structure and function of the isolated PAZ domain of eukaryotic AGOs was determined even before the first full-length prokaryotic structure was available. Several structural studies demonstrated simultaneously that the PAZ domain recognizes the 3′-end overhang of the guide RNA (Lingel et al., 2003; Ma et al., 2004; Song et al., 2003; Yan et al., 2003).

Studies on the 5′-end recognition revealed that the 5′-monophosphate of the guide RNA, which is generated during small RNA maturation, is required for the assembly of the small RNA into RISC (Chiu and Rana, 2002; Tomari et al., 2004). The domain responsible for 5′-end recognition in AGOs is termed the MID domain. Two independent structural studies characterized the binding pocket and hence how guide RNA-binding is achieved in prokaryotic AGOs. These studies also proposed a mechanism how distance-based measurement for target cleavage through the PIWI domain is achieved (the target is always cleaved opposite of nucleotide 10 and 11 of the guide strand) (Ma et al., 2005; Parker et al., 2005).

The PIWI domain has been shown to be the catalytic engine of the Argonaute protein and responsible for the endonucleolytic cleavage of the target RNA. Crystal structures revealed an unexpected homology to the RNase H family of enzymes (Parker et al., 2004; Song et al., 2004). Proteins belonging to the class of RNase H enzymes catalyze the hydrolysis of phosphodiester bonds in their respective DNA or RNA substrates. Thus the structural homology to RNase H, together with the conservation of critical aspartate residues important for catalysis, identified the PIWI domain being responsible for the slicer activity of the RNA-induced silencing complex. In addition the PIWI domain is also a binding partner for several AGO-interacting proteins (Friend et al., 2012). Nevertheless, the most well-characterized AGO-binding partners required for miRNA-mediated silencing of partially complementary targets are GW182 proteins, which are described in detail in the following section.

2.4.2 The GW182 protein family

In 2002, Eystathioy and colleagues discovered the GW182 protein (approx. 182 kDa in size with several glycine-tryptophan repeats) using sera from a patient with a motor and sensory polyneurophaty (Eystathioy et al., 2002). Three paralogues with overlapping function have been characterized in humans and vertebrates as well as various invertebrates (Trinucleotide repeat-containing gene 6 proteins; TNRC6A/B/C), whereas in insects only one paralogue is known so far (Ding and Han, 2007; Eulalio et al., 2009).

The domain architecture of the three human paralogues, TNRC6A, TNRC6B and TNRC6C, is shown in Figure 7. Subsequent studies revealed an essential role for GW182 protein family members in the miRNA-mediated gene silencing pathway (Ding et al., 2005; Jakymiw et al., 2005; Liu et al., 2005; Rehwinkel et al., 2005). It was demonstrated that GW182 proteins physically bind to a certain subset of Argonaute proteins and, more precisely, only to those proteins that are part of the miRISC (Ding et al., 2005; Liu

et al., 2005). Shortly afterwards, the N-terminal domain of GW182 was identified as being necessary and sufficient for binding the PIWI domain, with tryptophan residues located in the N-term being crucial for this interaction (El-Shami et al., 2007; Eulalio et al., 2009; Till and Ladurner, 2007; Till et al., 2007). Consequently, the N-terminal domain of GW182 proteins was called the AGO-binding domain or ABD (see Figure 7). Interestingly, some Argonaute proteins (e.g. *Hs* AGO2) can participate in both, the siRNA pathway where cleavage of the target RNA occurs and in the miRNA pathway, in which GW182 proteins are essential binding partners to mediate silencing. Like mentioned before, one major difference between siRNAs and miRNA in animals is that miRNAs are only partially complementary to their RNA targets (Braun et al., 2013; Fabian and Sonenberg, 2012; Huntzinger and Izaurralde, 2011). As a consequence, the targeted mRNA is not cleaved by the Argonaute protein. Nevertheless, how GW182 proteins can sense on a molecular basis whether the AGO protein is loaded with a miRNA (and thereby be a binding partner) or a siRNA has not yet been determined.

Recently it has been reported that GW182 proteins also directly associate with the two major cytosolic deadenylases PAN2-PAN3 and CCR4-NOT (Braun et al., 2011; Chekulaeva et al., 2011; Fabian et al., 2011) and hence being important for the recruitment of the major cytosolic deadenylases to miRNA targets. The binding to both deadenylase complexes is also mediated through tryptophan residues, but these tryptophan residues are present in the C-terminal silencing domain (SD) of GW182 proteins (Chekulaeva et al., 2011). Motifs critical for the binding to the CCR4-NOT complex have been identified and termed CIM motifs (CCR4-NOT interacting motifs) (Fabian et al., 2011).

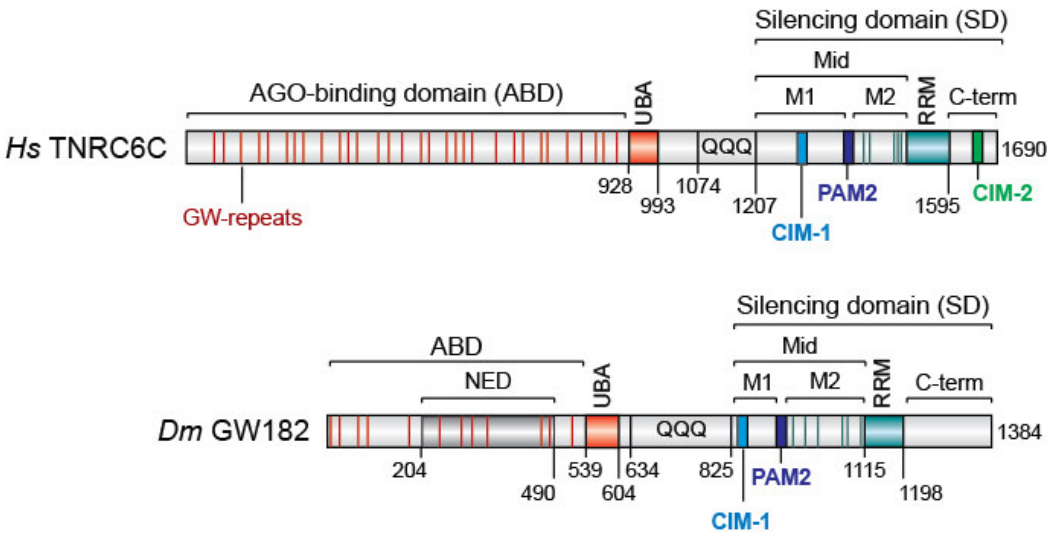


Figure 7: Domain organization of the human TNRC6C as well as the *Drosophila melanogaster* GW182 proteins (modified from Huntzinger et al. 2013). ABD, AGO-binding domain; UBA, ubiquitin associated-like domain; QQQ, region rich in glutamine; CIM-1 and CIM-2, CCR4-interacting motif 1 and 2; MID, middle region; PAM2, PABP-interacting motif 2; RRM, RNA recognition motif. Vertical red lines indicate the positions of GW-repeats crucial for AGO-binding. Vertical blue lines indicate the position of tryptophans important for NOT1 and PAN3 binding.

3 Motivation and aims

3.1 Structural and mechanistic insights into the cytoplasmic PAN2-PAN3 and CCR4-NOT deadenylase complexes

As mentioned in the introduction (see section 2.2), the PAN2-PAN3 and the CCR4-NOT complexes are the two major cytoplasmic deadenylase complexes in mammalian cells. These complexes commonly feature active deadenylase subunits bound to non-catalytic subunits that are essential for efficient deadenylation *in vivo*. The non-catalytic subunits often act as scaffold proteins that are likely required to stabilize the deadenylase subunits, i.e. the PAN3 and NOT1 subunits (Wahle and Winkler, 2013). Second, these non-catalytic subunits associate with RNA binding proteins and other pathway-specific proteins to recruit the deadenylases to certain mRNA targets.

As an example, the PAN3 adaptor protein interacts with PABPC1 through its PAM2 motif to guide the PAN2 deadenylase to mRNA targets (Mangus et al., 2004). The NANOS protein family is another well-known RNA-binding protein that has been shown to directly interact with the CNOT1 scaffold to promote deadenylation of specific mRNA targets particularly in the germline (Suzuki et al., 2010; Suzuki et al., 2012). Furthermore, the GW182/TNRC6 proteins associate with PAN3 and NOT1 to promote translational repression and mRNA decay of miRNA targets (Braun et al., 2011; Chekulaeva et al., 2011; Fabian et al., 2011; Huntzinger et al., 2013). However, the mechanism of how translational repression is induced by the CCR4-NOT complex remained enigmatic so far, although Meijer *et al*. recently proposed that eIF4A2 is recruited to mRNA targets through the CNOT1 MIF4G domain thereby imposing translational repression (Meijer et al., 2013). Interestingly the CNOT1 MIF4G domain has been shown in yeast to interact directly with another DEAD-box helicase protein, namely the translational repressor DDX6 (Maillet and Collart, 2002).

During my PhD I was aiming to elucidate how these two major cytoplasmic deadenylase complexes assemble on an atomic level, how they are recruited to miRNA targets, and, in the case of CCR4-NOT, induce translational repression. When I started these projects, structural data on the non-catalytic subunits of the PAN2-PAN3 and the CCR4-NOT complexes were not available, despite their importance for general mRNA degradation and the miRNA pathway.

The PAN3 adaptor protein consists of an N-terminal unstructured region that contains a PAM2 motif essential for PABPC1 binding, followed by a central domain predicted to adopt a pseudokinase fold and a highly conserved C-terminal domain of unknown three-dimensional structure. Self-interaction of the PAN3 protein, described earlier, could be attributed to the pseudokinase and the C-terminal domains. Additionally it was demonstrated that the C-terminal domain was sufficient to mediate binding to the PAN2 catalytic subunit (Mangus et al., 2004).

The CCR4-NOT complex consists of two conserved modules: a catalytic module that contains two deadenylases (CCR4a or CCR4b and CAF1 or POP2) and a NOT-module, minimally formed by the C-termini of NOT1, NOT2 and NOT3 (see section 2.2.3; Figure 4) (Bawankar et al., 2013). The C-terminus of NOT1 is highly conserved and defines the so-called NOT1 superfamily homology (SH) domain. The C-termini of NOT2 and NOT3 are highly related to each other and comprise the NOT-box domain (Zwartjes et al., 2004). Due to the lack of sequence homology to known structural folds it was not possible to predict the fold of the individual domains of NOT1, NOT2 and NOT3 and hence their assembly in the complex.

Finally it was shown that W-residues present in the C-terminal silencing domain of GW182 proteins (see Figure 7; blue bars) mediate the recruitment of deadenylation complexes to miRNA targets. Thus we sought to identify tryptophan binding pockets present in PAN3 and NOT1 to gain structural and mechanistic insights into this interaction.

3.2 How do eukaryotic Argonautes fulfill their function in small-RNA mediated gene silencing pathways?

By January 2009, the starting date of my PhD thesis, work from several groups had provided structural and mechanistic insights into the three-dimensional architecture of full-length prokaryotic AGOs, clarifying how these molecules bind nucleic acids (Ma et al., 2005; Parker et al., 2005; Song et al., 2004; Wang et al., 2008a; Wang et al., 2008b). Structural information on eukaryotic Argonaute proteins however, was limited to the isolated PAZ domain at that time (Lingel et al., 2003; Song et al., 2003; Yan et al., 2003).

Consequently, no structural information was available for the MID and the PIWI domains of eukaryotic AGOs, despite their importance in RNA binding and catalysis. Moreover, several independent studies focused on the fold of the MID domain and its role in miRNA-mediated gene silencing (Djuranovic et al., 2010; Kinch and Grishin, 2009; Kiriakidou et al., 2007). Based on the observation that human AGO2 binds to m^7GTP Sepharose beads it was proposed that certain eukaryotic AGO MID domains (in particular those AGOs which participate in the miRNA pathway) can bind the 5′-cap structure of target mRNAs to repress translation. Two different modes of cap-structure binding had been proposed: Early on, Kiriakidou and colleagues hypothesized that the MID domains of eukaryotic AGOs adopt an eIF4E-like fold, thus allowing the MID domain to sandwich the mRNA cap structure between two highly conserved phenylalanine residues in a mode similar to that observed in eIF4E (Marcotrigiano et al., 1997). In contrast, Djuranovic and co-workers identified a second, miRNA-dependent binding site located near the guide RNA binding pocket, which can bind nucleotides such as the 5′-cap structure. In this newly identified potential binding pocket, an aspartate residue was hypothesized to be central for allosteric control. To discriminate among the proposed models it was absolutely necessary to obtain structural information on the MID and PIWI domains at atomic resolution.

Furthermore, although the domain composition is conserved between prokaryotic and eukaryotic counterparts, the sequence conservation is low (less than 20%) and the eukaryotic proteins exhibit numerous insertions in their sequences. Therefore, it remained to be elucidated, whether guide RNA binding in eukaryotic AGOs resembles the mode of DNA binding observed in prokaryotic proteins and whether the aforementioned insertions are important for nucleic acid binding or the recruitment of eukaryotic specific binding partners like the GW182 proteins.

In essence, structural and functional analyses on the MID and the PIWI domains were crucial for improving our knowledge of eukaryotic Argonaute biology, especially since the PAZ domain was the only determined domain of eukaryotic Argonautes at that time.

4 Results and discussion

My PhD work was aimed at providing a detailed molecular understanding of general and specific posttranscriptional gene regulation mechanisms. To this end, I solved the crystal structures of protein subunits or domains required for the assembly of the two major deadenylase complexes PAN2-PAN3 and CCR4-NOT, as well as of a key player in silencing pathways, the Argonaute protein QDE-2 from *Neurospora crassa*. During this work evidence emerged, that these two processes are directly interlinked in such a way, that AGOs acting in the miRNA silencing pathway, recruit the general deadenylase complexes to miRNA-targets through a bridging factor termed GW182. The assembly of this complex triggers translational repression and mRNA degradation. In the following two chapters I will summarize the results I obtained during my PhD, starting with the deadenylase complexes.

4.1 The structure of the PAN3 pseudokinase and its interaction with GW182 proteins

All experimental data and detailed descriptions about the experimental procedures can be found in the attached manuscript (Christie et al., 2013).

Sequence analysis as well as secondary and tertiary structure prediction tools like PSIPRED (Buchan et al., 2013) and HHPRED (Soding et al., 2005) demonstrated that the PAN3 adaptor protein can be divided into three distinct regions: An N-terminal unstructured region (N-term) which mediates binding to PABC1 through a PAM2 motif, a central pseudokinase domain (PK), which is predicted to be inactive and a C-terminal conserved domain (C-term), which has been shown to be necessary and sufficient to mediate binding to the catalytic subunit PAN2 (Mangus et al., 2004) (see Figure 3). Due to the lack of sequence homology to known structural folds however, the structure of the C-terminal conserved domain remained obscure. In addition, it was known that

PAN3 can self-interact, however the stoichiometry of this self-interaction was not defined.

By utilizing structure prediction tools I designed protein fragments of *Nc* PAN3 suitable for protein expression and ultimately for protein crystallization. I optimized the purification conditions for several constructs from different species and demonstrated by thermal shift assays (Lucet et al., 2013) that PAN3 proteins are stabilized upon ATP-binding. Interestingly, human PAN3 protein fragments could only be purified when nucleotides were supplemented during the purification process. Furthermore crystals of PAN3 were only obtained in the presence of nonhydrolyzable ATP analogues and $MgCl_2$ in the protein solution (ATP-γ-S in the case of *Nc* Pan3 and AMPPNP for *Dm* PAN3).

In a shared project, we solved the crystal structures of the folded domains of the wild-type and a mutant version of *Nc* PAN3, as well as *Dm* PAN3 at 3.3, 2.85 and 3.6 Å resolution, respectively (Christie et al., 2013). Using a *Nc* PAN3 protein fragment I obtained the first crystals of a PAN3 protein and enhanced the crystal quality by using different optimization methods like microseeding. Thereby I improved the resolution of the wild-type protein to 3.3 Å from initially only poorly diffracting crystals (less than 4.5 Å). I determined initial phases by the single-wavelength anomalous diffraction (SAD) method (Rice et al., 2000) using a selenomethionine substituted version of the protein. Phase determination was sufficient to build an initial model that was then used to solve and refine the structural model of the *Dm* PAN3 protein that had been crystallized by my colleague Tara Christie. In a bid to improve the resolution of our structural models, I designed and crystallized a mutant version of *Nc* PAN3 that indeed diffracted X-rays up to 2.85 Å and thus strengthened the conclusions that we made from the models.

To determine the stoichiometry of PAN3 oligomers in solution, we performed size exclusion chromatography runs coupled to multiangle static laser-light scattering (MALLS) (Folta-Stogniew and Williams, 1999). These experiments indicated a molecular

weight of approximately 100 kDa for the *Nc*, *Dm* and *Hs* PAN3 proteins, in contrast to the expected size of roughly 50 kDa for a monomeric PAN3. This indicates that PAN3 forms dimers in solution, and that dimer formation is conserved throughout evolution. In contrast, the PK and the C-term domains of *Dm* PAN3 in isolation behaved like monomers in solution, showing that dimer formation only occurs, when the protein fragment contains both folded domains.

PAN3 not only forms dimers in solution, but also crystallizes as a homodimer. This homodimerization is mainly mediated via a central coiled coil domain that links the N-terminal pseudokinase domain (PK) to the C-terminal knob (CK) domain (see Figure 8). The central left-handed coiled coil domain is asymmetric due to a kink in only one α-helix of the coiled coil domain. The N-terminal pseudokinase domain displays, as expected, a canonical bilobal kinase domain fold (Bryant et al., 1974). Nevertheless the PK domain of PAN3 exhibits numerous substitutions in almost all conserved motifs that are crucial for kinase activity, i.e. the catalytic VAIK and HRD motifs. Another striking difference between PAN3 and canonical kinase domains is the lack of an activation loop, which usually bridges the DFG motif and the APE motif (Hanks et al., 1988). Therefore I tested whether PAN3 retained the ability to hydrolyze ATP *in vitro*. As expected, I could not detect any intrinsic ATPase activity, even after a prolonged incubation time of 20 hours.

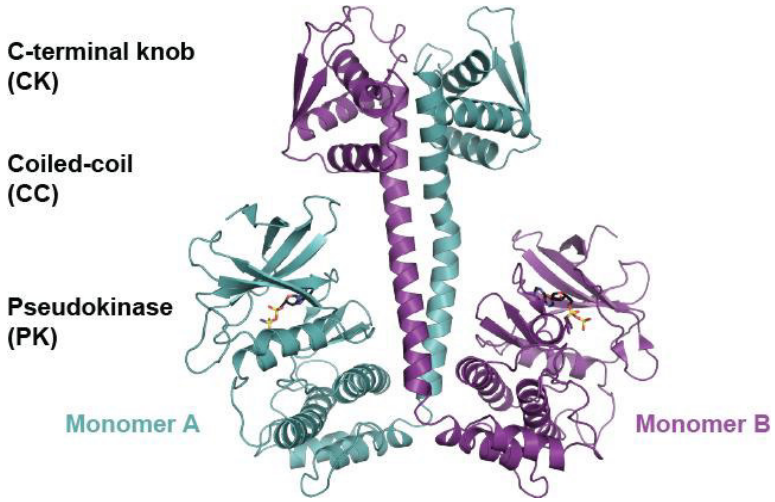


Figure 8: Crystal structure of the *Dm* PAN3 pseudokinase dimer in complex with AMPPNP (adapted from Christie *et al*. 2013). The PAN3 monomers are colored in cyan and purple respectively.

The CK domain adopts a compact fold consisting of 3 α-helices (helix α2, α3 and α4) with a β-hairpin inserted between helix α3 and α4. Comparison of the CK domain to all deposited structures in the PDB database using the DALI-server revealed that the CK domain adopts a rather unique structural fold. With the structure in hand, we next analyzed the surface conservation of the CK domain to identify highly conserved and surface-exposed residues that could mediate PAN2 binding. Coimmunoprecipitation assays performed by my colleague Eric Huntzinger helped us to identify a number of surface-exposed residues involved in PAN2 binding (see Figure 5 **C** and **D** in Christie et al., 2013).

Notably, crystal packing interactions within the *Dm* PAN3 crystal lattice unveiled the presence of a tryptophan-binding pocket at the base of the coiled coil domain (a neighboring symmetry mate in the crystal inserts a W-residue into this pocket). This pocket is highly symmetrical and formed by both monomers. It is lined by several hydrophobic residues including M651, G655 and F658. Importantly proline 652 in each monomer serves as a stacking partner for the bound tryptophan residue (W707). Consequently, this hydrophobic pocket is formed upon *Dm* PAN3 dimerization (see Figure 9). Again, mutagenesis of residues lining this W-binding pocket or mutants that abolish *Dm* PAN3 homodimerization efficiently abolish TNRC6C binding *in vitro* and *in vivo*, thus validating the importance of this W-binding pocket for TNRC6 binding and hence for the recruitment to miRNA targets (see Figure 6 **D** and **E** in Christie et al., 2013). By sequence comparisons using Jalview (Waterhouse et al., 2009), I could show that the presence of the tryptophan-binding pocket correlates well with the presence of TNRC6/GW182 proteins in higher eukaryotes.

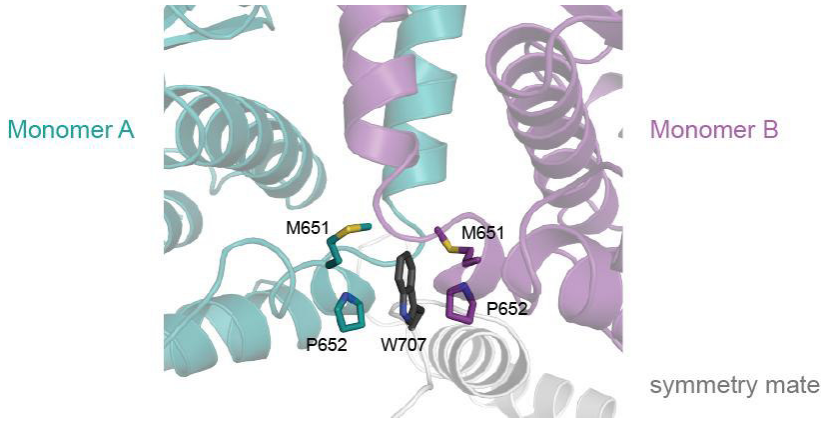


Figure 9: Close-up view of the W-binding pocket (adapted from Christie *et al*. 2013). The two monomers forming a dimer are shown in cyan (Monomer A) and purple (Monomer B) respectively. A neighboring symmetry mate (gray) inserts an exposed W-residue (black) in a hydrophobic pocket at the base of the coiled coil domain. Selected residues forming the W-binding pocket are shown as sticks.

In summary, these structures provides a detailed understanding of the assembly of the PAN2-PAN3 complex, the self-dimerization of PAN3 and most importantly, the recruitment of the complex to miRNA targets through GW182/TNRC6 proteins.

4.2 The CCR4-NOT complex

To gain functional and structural insights into the mechanisms of deadenylation and translational repression carried out by the CCR-NOT complex, I determined several sub-complexes of the CCR4-NOT complex in two collaborative projects.

In a first project, we initially solved the structures of the NOT-box domains of the human CNOT2 and CNOT3 proteins in isolation (both at 2.4 Å resolution), as well as the superfamily homology domain (SHD) of NOT1 from *Chaetomium thermophilum* (at 3.2 Å resolution). The highly conserved NOT-box domain is shared by NOT2 and NOT3 and located at their respective C-termini. Similarly, the C-terminus of NOT1 is highly conserved and defines the superfamily homology domain. In a previous study from the lab, it was demonstrated that the C-termini of NOT1, NOT2 and NOT3 form a complex,

the so called NOT-module (Bawankar et al., 2013). Next we also determined the structure of the assembled NOT-module core at 3.2 Å resolution.

In an independent second project, we investigated the structural basis of the interaction between TNRC6/GW182 proteins and the CCR4-NOT complex and thus how the deadenylases are recruited to miRNA targets. We showed that a DUF3819 domain (domain of unknown function 3819) in CNOT1 is the binding site for the highly conserved subunit CNOT9 and that the binary complex consisting of CNOT1 and CNOT9 provides a major binding site for GW182 proteins. We determined the structure of this sub-complex at 1.65 Å resolution without and at 2.05 Å resolution in complex with tryptophan (free L-tryptophan was supplemented to the crystallization condition).

Next we structurally addressed the question, how the CCR4-NOT complex could achieve its repressive activity on translation. As described previously (see section 3.1), it was known that NOT1 can bind the translational repressor and decapping activator DDX6 (Maillet and Collart, 2002) in yeast and in fly embryonic extract (Temme et al., 2010). We determined the domain in CNOT1 responsible for binding to the DDX6 RecA-C domain and solved this sub-complex at 1.75 Å resolution. This complex links decapping to deadenylation through direct interactions.

4.2.1 The crystal structure of the human NOT-module core (CNOT1/2/3)

All experimental data and detailed descriptions about the experimental procedures can be found in the attached manuscript (Boland et al., 2013).

Again, by using secondary structure prediction and sequence alignment tools I designed all constructs used for crystallization purposes in this study and designed specific mutants for functional studies that abrogate NOT-module formation. I optimized the protein purification procedure and determined the optimal protein buffer conditions for crystallization by using thermal shifts assays (Lucet et al., 2013). In a first

attempt we crystallized the individual domains of the C-terminus of NOT1 from *Chaetomium thermophilum* and the human NOT-box domains of CNOT2 and CNOT3. I worked on all critical steps during the structure determination process like data collection, phase determination and structure refinement. Notably, all the isolated domains were solved by experimental phasing which paved the way to ultimately solve the structure of the ternary CNOT1/2/3 complex by molecular replacement.

The C-term of CNOT1 consists of two tightly interacting subdomains that are arranged as two perpendicular stacks of HEAT-like repeats. The N-terminal subdomain comprises six α-hairpin repeats, the C-terminal subdomain consists of four HEAT-like repeats. These subdomains are connected by a long surface-attached linker.

Next we showed by multi-angle laser light scattering and size exclusion chromatography that a protein fragment encompassing the CNOT3 NOT-box domain dimerizes in solution. CNOT3 also forms dimers in the crystal environment. The NOT-box domain contains three α-helices followed by a five-stranded half-open β-barrel, similar to the SH3-type domain. In contrast to the SH3-fold, the NOT-box of CNOT3 homodimerizes via its tightly packed N-terminal α-helix bundle. The interface exhibits a two-fold symmetry and is mainly composed of hydrophobic residues (see Figure 11).

In contrast, CNOT2 crystallizes as a domain-swapped dimer in which the N-terminal helices adopt an open conformation. CNOT2 packs as a dimer of dimers in the crystal leading to a fairly symmetrical homotetramer. The formation of dimers and tetramers can also be observed in solution measured by MALLS experiments. To discriminate between these two different binding modes, in the context of the assembled NOT-module, we crystallized a ternary complex additionally containing CNOT1.

In a bid of crystallizing a ternary complex, I designed crystallization constructs compatible for co-expression in *Escherichia coli* cells (Diebold et al., 2011). In the context of the ternary complex, N-terminal sequences upstream of the NOT-box domains were included in the expression constructs of CNOT2 and CNOT3 because these regions were

shown to be crucial for complex assembly (Bawankar et al., 2013). Importantly, these upstream sequences were predicted to be devoid of any secondary structure element. I optimized the purification process of the ternary complex and used the purified complex for biochemical characterization. I identified a stable NOT-module core by performing limited proteolysis experiments with various proteases. Notably, solely the CNOT1 scaffold protein was degraded in these experiments. In contrast, the CNOT2 and CNOT3 remained intact despite the presence of the N-terminal upstream sequences that were predicted to be unstructured. In a collaborative effort with my colleagues, the structure could be solved at 3.2 Å resolution.

The heterotrimeric complex consisting of CNOT1, CNOT2 and CNOT3 displays a 1:1:1 stoichiometry and is organized into a symmetrical lobe containing the NOT-boxes of CNOT2 and CNOT3 and an asymmetrical lobe containing the CNOT1 scaffold protein (see Figure 10).

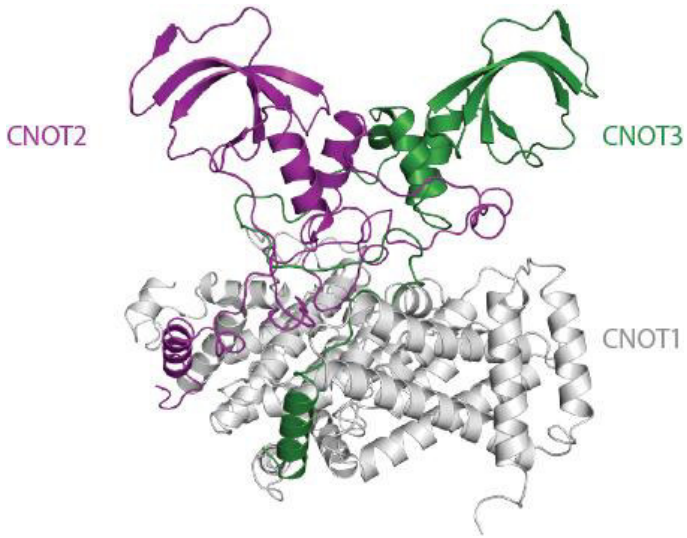


Figure 10: Crystal structure of the minimal NOT-module core showing CNOT1, CNOT2 and CNOT3 in gray, purple and green respectively (adapted from Boland *et al*. 2013).

The structures of CNOT1 and CNOT2 in the complex are virtually identical to the isolated domains, whereas CNOT2 displays a closed conformation of the N-terminal α-helices when crystallized as a ternary complex. As a consequence, CNOT2 and CNOT3 form a highly symmetrical heterodimer, similar to the CNOT3 homodimer. Importantly,

the heterodimerization interface shows a higher degree of complementarity regarding the side chain interactions than the homodimerization interface does, which explains in part why heterodimerization is favored in solution (see Figure 11).

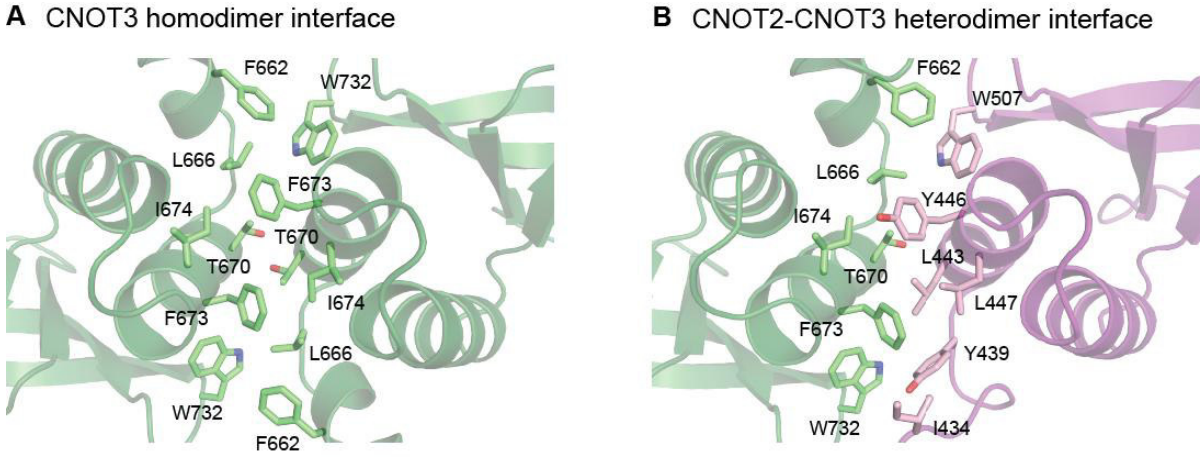


Figure 11: Comparison between the CNOT3-CNOT3 homodimer interface and the CNOT2-CNOT3 heterodimer interface (adapted from Boland *et al*. 2013). The CNOT3 and CNOT2 monomers are shown in green and in purple, respectively. Selected side chains forming the hydrophobic interfaces are shown in lime green and purple for CNOT3 and CNOT2, respectively.

The N-terminal sequences upstream of the NOT-box domains wrap around the respective opposed NOT-box domain like clamps and lock themselves back through side chain interactions. The heterodimer is bound onto the NOT1 surface by further N-terminal sequences, each of which consists of an N-terminal helix, followed by a region entirely devoid of any secondary structure. Those N-terminal helices insert into lateral grooves on the NOT1 surface (see Figure 10). The adjacent unstructured regions mold onto surface grooves on the NOT1 surface thereby tethering the symmetric lobe onto the asymmetric lobe. Lastly, the C-terminal tail of CNOT2 clamps the N-terminal residues of CNOT3 onto the CNOT1 surface, thereby preventing dissociation of CNOT3. This mode of assembly of the CCR4-NOT complex, mediated by intrinsically disordered regions (IDRs) could not have been detected without determining the crystal structure of the minimal NOT-module core.

Mutagenesis studies done by my colleague Dyugu Kuzuoğlu-Öztürk surprisingly revealed that substitutions of only two or three surface residues on CNOT1 that directly contact the IDRs of CNOT2 or CNOT3, are sufficient to prevent binding. Similarly, deletion mutants of the IDRs of CNOT2 and of CNOT3 prevent binding to CNOT1 and the respective opposed NOT-box protein.

In summary, the NOT-module core structure was solved in a two step approach and the importance of the NOT-module for correct complex assembly and function was demonstrated in functional assays, i.e. complementation assays (see Figures 6 and 7 in Boland et al., 2013).

4.2.2 Recruitment of the CCR4-NOT complex by GW182 proteins

All experimental data and detailed descriptions about the experimental procedures can be found in the attached manuscript (Chen, 2014).

Initial domain mapping experiments done by coimmunoprecipitation assays, revealed that GW182/TNRC6 proteins mainly interact with the DUF3819 domain present in the MID region of CNOT1 (see Figure 4). This domain of CNOT1 was also shown to mediate binding to the subunit CAF40/CNOT9 and we therefore termed it CN9BD. To gain structural insights into this dual function of the CNOT1 CN9BD domain, we crystallized a complex containing the CN9BD domain and CNOT9 in both the absence and the presence of free L-tryptophan in the crystallization condition. The crystals allowed us to determine the structures at 1.65 Å without and at 2.05 Å with bound tryptophan.

Similarly to the described PAN3 and NOT-module projects before, I initiated and established the project in the lab. I designed all, and generated most constructs used in this project for protein expression and crystallization of the CN9BD, as well as constructs of CNOT9 suitable for co-expression with the CN9BD. In a crucial experiment

I showed that the CN9BD in isolation is unstable but gets highly stabilized when bound to the conserved ARM domain of the core subunit CNOT9. I established the conditions for complex purification and purified in an collaborative effort several different complexes which I set up for crystallization trials and used for *in vitro* pull-down assays. Furthermore, I worked on all critical steps during the structure determination process including crystal optimization as well as data collection, structure solution and refinement.

The crystal structure of the highly conserved CNOT9 subunit was described in 2007 (Garces et al., 2007) and used as a molecular replacement model to determine the structure of the CN9BD-CNOT9 complex. The CNOT9 subunit in isolation is virtually identical to CNOT9 in complex with CNOT1, as is indicated by a low root mean square deviation (r.m.s.d.). Six imperfect Armadillo (ARM) repeats are formed by 17 α-helices that are arranged in a clockwise spiraling manner. The CN9BD domain consists of 7 α-helices in which helices α2–α4 are arranged as an antiparallel, rod-shaped bundle of three kinked α-helices (see Figure 12).

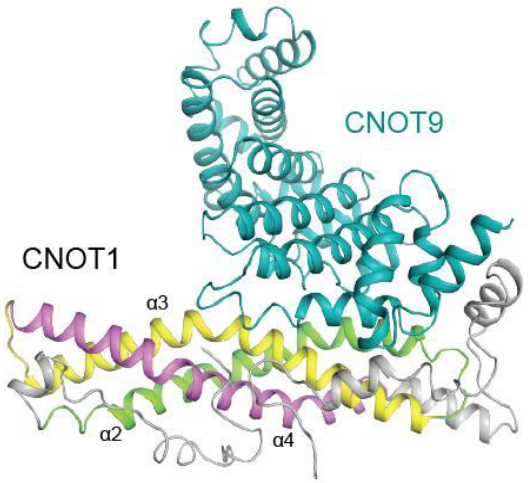


Figure 12: Overview of the CNOT1-CNOT9 binary complex showing CNOT9 in cyan and CNOT1 helices α2, α3 and α4 in green, yellow and purple, respectively (adapted from Chen *et al*. 2014). The W-binding pockets present in CNOT9 are not shown.

CNOT9 forms homodimers in solution. This homodimerization interface largely overlaps with the heterodimerization interface with CNOT1. The interface between

CNOT1 and CNOT9 is larger than the homodimerization interface (3170 Å^2 vs 2005 Å^2) and hence likely to be favored. The interaction between the CNOT1 CN9BD and CNOT9 ARM domains is mainly mediated via hydrophobic interactions and assisted by only a few hydrogen bonds. To validate that the interface observed in the crystal environment also forms in solution, we designed CN9BD and CNOT9 mutants and performed *in vitro* pull-downs as well as coimmunoprecipitation experiments in human and fly cells. These experiments revealed that single point mutations reduced but did not abrogate the interaction between CN9BD and CNOT9. To disrupt binding of the CNOT9 to CNOT1 it was necessary to introduce four point mutations in CNOT9. Conversely, to disrupt binding of CNOT1 to CNOT9, it was necessary to introduce five specific point mutations. These results are consistent with a rather extensive and high affinity interaction, as observed in the crystal structure.

However the most important observation was the presence of tryptophan residues inserted into tandem tryptophan pockets located on the convex surface of the horseshoe like ARM-domain of CNOT9 (see Figure 13).

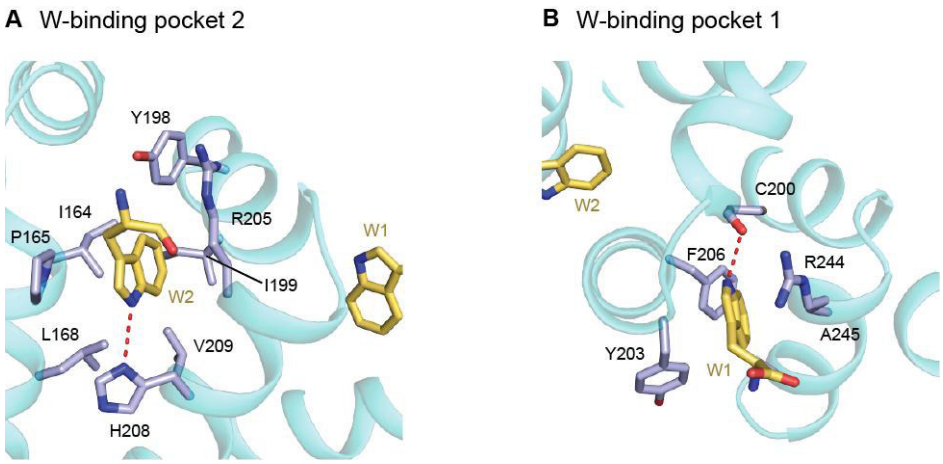


Figure 13: Close up view of the tandem tryptophan binding pockets identified in CNOT9. Selected side chains surrounding the tryptophan residues (yellow) are shown as sticks colored in purple. Tryptophan 2 (W2) is only present when L-tryptophan is supplemented in the crystallization condition (**A**). Tryptophan 1 (**B**) is presumably part of a disordered C-terminal tail from a neighboring symmetry mate within the crystal lattice (adapted from Chen *et al*. 2014).

To elucidate whether these W-binding pockets are important for binding to GW182/TNRC6 proteins I performed *in vitro* pulldown asssays. The results have been confirmed in cells as well. I mutated residues either lining W-binding pocket 1 or pocket 2 or I combined those mutations. The pulldowns showed that mutation in either one of the two pockets reduced but did not completely abolish TNRC6 binding (see Figure 14 lanes 8 and 9), whereas combining those mutations efficiently abrogated binding (see Figure 14 lanes 10-12). Importantly, a complex containing a mutated version of CNOT9 bound to the CN9BD of CNOT1 did not interact with TNRC6A *in vitro*, indicating that tandem W-binding pockets in CNOT9 represent the major binding pockets for W-rich proteins (see Chen et al., 2014 Figure S1**B**).

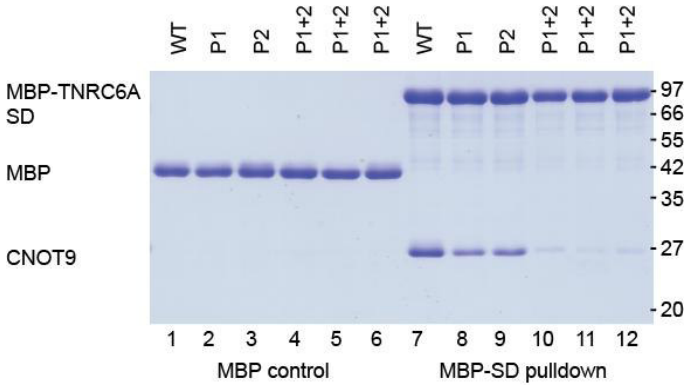


Figure 14: *In vitro* pull-down analysis between the silencing domain (SD) of TNRC6A and the ARM domain of CNOT9, using either wildtype or mutated pocket mutants of CNOT9. Pocket 1 (P1), pocket 2 (P2) and a combination of various mutations in pocket 1 and pocket 2 (P1+2) (Chen *et al*. 2014).

4.2.3 A complex comprising CNOT1 and the translational repressor and decapping activator DDX6 links deadenylation to decapping

In a second part of this project, we tested a hypothesis previously proposed (Meijer et al., 2013) – namely, that the central MIF4G domain of CNOT1 interacts with eIF4A2 in an eIF4G-eIF4A1-like manner to interfere with translation. To this aim, coimmunoprecipitation experiments in human HEK293 cells were performed by my colleagues. In these experiments, we included another helicase, termed DDX6, as a positive control, because this helicase was previously described as an interaction partner

to the N-terminal half of NOT1 in yeast (Maillet and Collart, 2002). In contrast to Meijer and co-workers, we did not observe binding of DDX6 to CNOT1, but we clearly detected binding of DDX6 to the CNOT1 MIF4G domain. In a subsequent step, we determined the crystal structure of the CNOT1 MIF4G domain in complex with the RecA-C domain of DDX6.

The protein complex was co-expressed in *E.coli* star cells using two previously described plasmids (Petit et al., 2012; Tritschler et al., 2009). Although the crystals tend to grow in clusters, I mechanistically separated a few single plates which finally led to a 1.75 Å data set. We used the previously determined crystal structures of the DDX6 RecA-C domain and the CNOT1 MIF4G domain as molecular replacement models. Similarly to the CNOT1-CNOT9 complex, I contributed to every step during structure determination, including phase determination by molecular replacement and refinement, as well as I analyzed the structure and designed binding mutants.

In sharp contrast to the CN9BD-CNOT9 complex, the interface between the CNOT1 MIF4G domain and the RecA-C domain of DDX6 is rather small (approximately 640 Å) and is dominated by polar interactions (see Figure 15). This might explain why DDX6 is not purified as a constitutive component of the CCR4-NOT complex in diverse studies (Bai et al., 1999; Gavin et al., 2002; Liu et al., 1998; Morel et al., 2003), although Me31b (the *Drosophila melanogaster* orthologue of DDX6) was found to coimmunoprecipitate with NOT1 using a monoclonal NOT1 antibody in *Drosophila* embryo extract (Temme et al., 2010). The yeast orthologue DHH1 was also described as interaction partner of NOT1 (Maillet and Collart, 2002).

Two structural interface elements present in DDX6 are of particular interest. Firstly, a highly conserved Arginine residue (R375) in DDX6 serves as an “anchor” residue by inserting deeply into a highly conserved patch of CNOT1, thus forming multiple interactions, incuding stacking interactions and hydrogen bonds (see Figure 15A). Secondly, loop L3 of DDX6 forms several hydrogen bonds to distinct side chains

of CNOT1, thereby providing specificity to this interaction (see Figure 15B). Hence we termed this loop L3 the specificity loop.

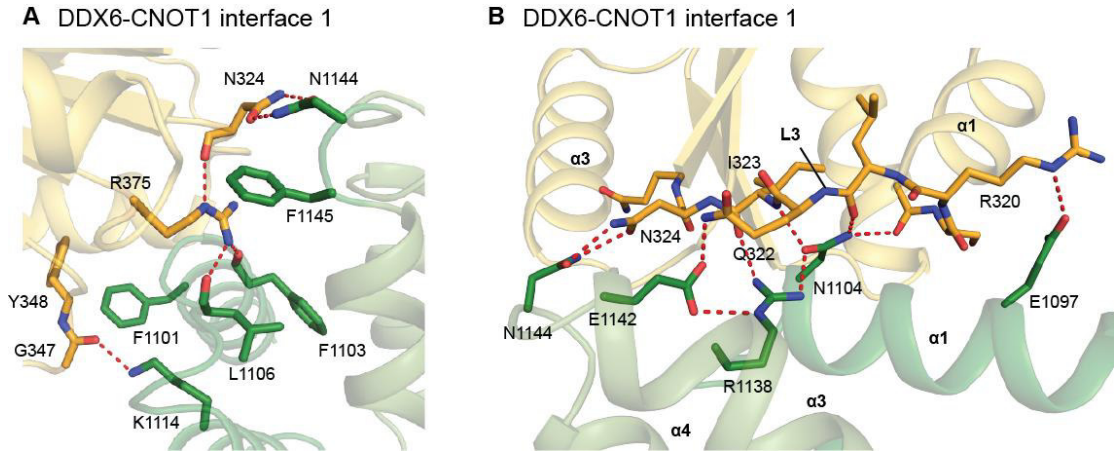


Figure 15: Close up-views showing the hydrophilic interface between the RecA-C domain of *Hs* DDX6 shown in yelloworange and human CNOT1 MIF4G domain in green. Selected residues are shown as sticks in orange and dark green for DDX6 and CNOT1, respectively. A highly conserved Arginine residue (R375) from DDX6 inserts into a conserved patch in CNOT1 (**A**). Specific site chain recognitions between Loop L3 from DDX6 and the first two HEAT repeats of CNOT1 (**B**) (Chen et al., 2014).

Next, we identified key residues important for complex formation and validated those by pulldowns or coimmunoprecipitation experiments.

In summary, in this project we showed that a binary complex consisting of the CNOT1 CN9BD and the CNOT9 ARM domains, assist the recruitment of the CCR4-NOT complex to miRNA targets through tandem W-binding pockets located in the highly conserved subunit CNOT9. In a second part we showed that the CNOT1 MIF4G domain and DDX6 form a complex thereby linking translational repression and decapping to deadenylation.

4.3 Structure and ligand binding of a eukaryotic Argonaute protein

To address the questions raised in section 3.2 on the fold and the biology of eukaryotic AGOs, I determined the crystal structures of both, the MID domain in isolation (Boland et al., 2010; Faehnle and Joshua-Tor, 2010) as well as the first MID-PIWI lobe (Boland et

al., 2011) of the eukaryotic AGO QDE-2 from the filamentous fungus *Neurospora crassa*. Experimental data and detailed descriptions about the experimental procedures can be found in the attached manuscripts (Boland et al., 2011; Boland et al., 2010).

In this chapter, two related projects are summarized. In both projects I conducted all *in vitro* experiments, including construct design, protein purification and crystallization as well as RNA binding assays.

The structure of the MID domain in isolation revealed, that the *Nc* QDE-2 MID domain adopts a Rossmann-like fold and thus ruled out the possibility that eukaryotic AGO MID domains adopt an eIF4E-like fold to sequester the mRNA cap structure between two conserved phenylalanine residues. Additionally, the structure of the human AGO2 MID domain was published simultaneously to the *Nc* MID domain (Faehnle and Joshua-Tor, 2010; Frank et al., 2010), confirming that MID domains in eukaryotic AGOs adopt a Rossmann-like fold. The QDE-2 MID domain superposes very well onto the previously determined MID domains of prokaryotic Argonaute proteins, with an r.m.s.d. value of 2.6 Å between the *Nc* MID and the *Af* MID domain, despite poor sequence conservation (9 %).

The 5′-guide RNA binding pocket however, is lined by several conserved residues mediating RNA binding. Among those highly conserved residues, Tyrosine 595 in *Nc* QDE-2 serves as a stacking platform for the 5′-terminal nucleotide of the guide RNA, which shows a strong bias towards uridine (Ghildiyal et al., 2008; Ghildiyal et al., 2010; Hu et al., 2009; Lau et al., 2001). The invariant K599 and K638 residues as well as the poorly conserved K634 coordinate a sulfate ion (S1) stemming from the crystallization condition (see Figure 16A). By superposing the *Af* MID-PIWI domains in complex with guide-RNA (Parker et al., 2005) (see figure 16 B) onto the *Nc* MID domain, the 5′-uridine base is positioned very well into the positively charged binding pocket (Figure 16 C). Importantly the 5′-phosphate of the guide RNA superposes almost

perfectly onto the sulfate ion S1. This model further validates Y595 as a stacking partner for the 5′-uridine base of the guide RNA.

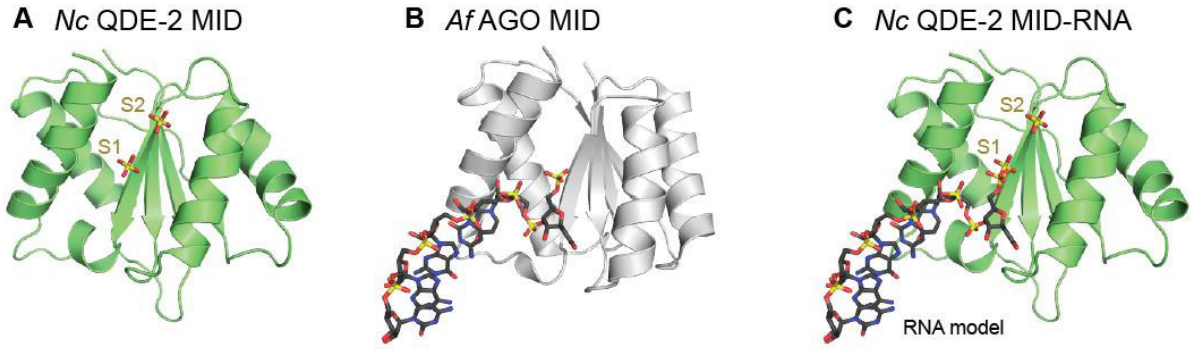


Figure 16: Comparison of the MID domain shown in lime from *Neurospora crassa* (**A**) and *Archaeoglobus fulgidus* bound to RNA, shown in gray (**B**). Modeling of an RNA substrate onto the *Nc* QDE-2 MID domain (**C**).

Surprisingly, the structure of the *Nc* QDE-2 MID domain revealed a second bound sulfate ion. This second binding site is located in close proximity to the first ligand binding site (distance between the two sulfate ions is ~6.3 Å), sharing two highly conserved lysine residues (K599 and K638) in common. As a result, these two lysines coordinate both sulfate ions simultaneously. Moreover, the proposed aspartate residue (D603 in QDE-2) implicated in allosteric control in a subset of eukaryotic AGOs (Djuranovic et al., 2010), is distant less than 7 Å from the second sulfate ion (S2). Thus, this second sulfate could indeed represent a second ligand that is bound in close proximity to the 5′-nucleotide binding site. Both binding pockets are not independent of each other, since the lysine residues K599 and K638 participate in binding of both ligands and might therefore provide an explanation for the allosteric effects observed by Djuranovic and co-workers. However whether this second ligand binding site would be accessible in the context of a full-length eukaryotic Argonaute remained still obscure. In fact evidence from the prokaryotic structures would occlude such a model since the very C-terminal tail of the AGO is deeply inserted into the MID-PIWI interface and would therefore not allow binding of a second ligand in this conformation.

In a second project I crystallized the first MID-PIWI lobe of a eukaryotic Argonaute protein. Despite numerous attempts to optimize the crystal quality and an extensive crystal screening process, the resolution of the obtained data sets from the wild-type protein fragment was restricted to 3.65 Å. In a two step approach, I then first determined and refined the crystal structure of the wild-type protein at low resolution and subsequently designed a deletion mutant of an internal loop element that was disordered in the crystal lattice. This deletion mutant crystallized in a different space group and I obtained crystals that diffracted X-rays to 1.85 Å resolution.

Three main conclusion could be drawn from this structure: First, not only the individual MID and PIWI domains superpose well with the previously determined archeal and eubacterial structures, but also the relative orientation of these two domains with respect to each other is highly conserved. Notably the C-terminal residues of the protein are deeply inserted into the MID-PIWI interface, reminiscent of the prokaryotic AGOs.

Second, by modeling a double-stranded nucleic acid substrate stemming from a *Tt* AGO-nucleic acid complex (Wang et al., 2009) (PDB ID code 3HJF) onto the *Nc* QDE-2 MID-PIWI lobe, I was able to identify a number of eukaryotic loops and their potential contribution to nucleic acid-binding. Briefly, I was able to show that loop L1 of the *Nc* QDE-2 MID-PIWI lobe is flexible and can adopt different conformations, probably depending on the bound nucleic acid substrate. In cases where the downstream duplex is formed, loop L2 is likely to fix the phosphodiester backbone of the target strand as well as nucleotide 14 from the guide strand, thereby supporting the positioning of the target strand onto the catalytic site. Loop L4 is disordered but based on the prokaryotic structures, this loop likely helps to orient and fix nucleotides 2-5 of the guide RNA, a region known as seed sequence. Finally, a striking feature of eukaryotic AGOs is a C-terminal insertion that in the case of QDE-2 folds into two tightly packed helices (see Figure 17). This insertion is also predicted to be helical in the human AGO proteins. By

superposing the full-length *Tt* AGO onto the QDE-2 MID-PIWI lobe, loop L1 and the C-terminal insertion of *Nc* QDE-2 end up in close proximity to the PAZ domain of the *Tt* AGO, but only when a miRNA is bound. This suggests, that loop L1 and the α-helical insertion establish conformation-dependent interaction with the PAZ domain in eukaryotic AGOs and thus these elements might act as sensors for the functional state of the AGO.

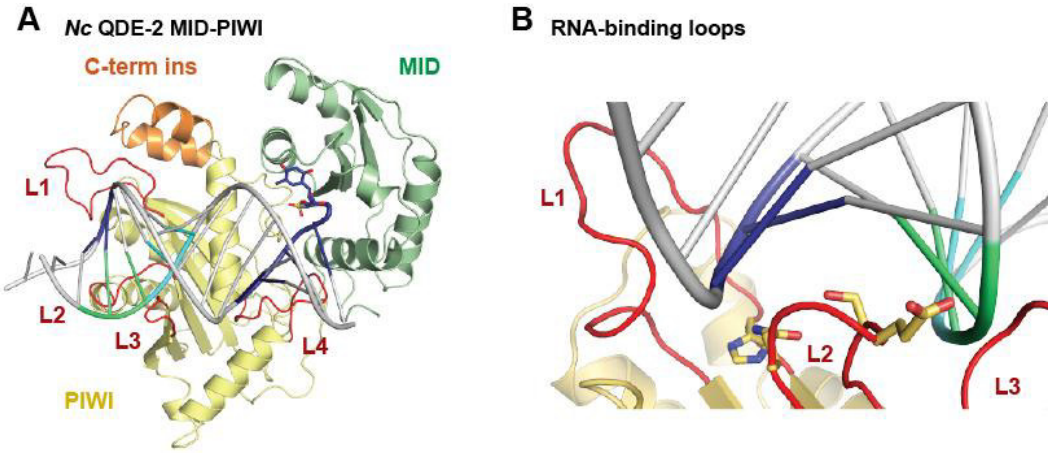


Figure 17: Crystal structure of the *Nc* QDE-2 MID-PIWI lobe. A double stranded nucleic acid from *Tt* AGO (PDB code 3HJF) is modeled onto the *Nc* QDE-2 MID-PIWI lobe (**A**). Bases that are recognized by RNA binding loops (red) are labeled in blue, green and cyan (**B**).

Third, as in prokaryotic AGOs, the 5′-guide RNA binding pockets forms at the MID-PIWI interface and is completed by its C-terminal tail. Consequently, the C-terminal residues are deeply inserted into the MID-PIWI interface thereby displacing the second sulfate ion (S2), observed in the crystal structure of the isolated MID domain. I was able to identify eukaryotic-specific residues important for the formation of the MID-PIWI interface. Among those are R895 and H899 in the PIWI domain and D603 in the MID domain. D603 forms hydrogen bonds to R895 and H899. RNA binding studies demonstrated that mutations in any of these residues leads to the destabilization of the MID-PIWI interface, thereby reducing or abolishing RNA binding without having to invoke allosteric regulation (see Figure 18).

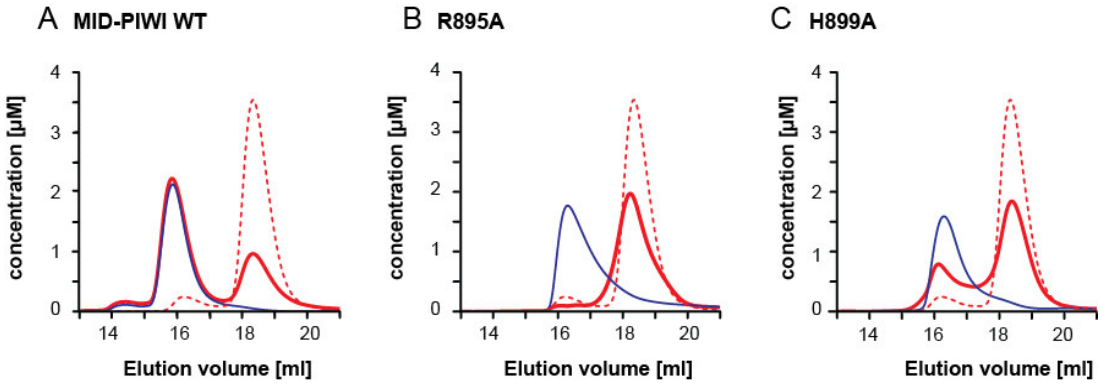


Figure 18: RNA binding studies using a gel filtration based assay (Muller et al., 2006). Free QDE-2 MID-PIWI protein is shown as blue solid line, free RNA as dashed red line and a mixture is shown as red solid line. A shift of RNA upon mixing with the QDE-2 protein from low to high molecular weight demonstrates complex formation. RNA binding is clearly shown in a 1:1 stoichiometry for the wild-type protein fragment (**A**), whereas single point mutants show a reduced nucleic acid binding affinity (**B** and **C**).

5 Conclusions

In summary, this thesis can be divided into two distinct but interlinked parts. In a chronological order, I was first working on the structure and function of eukaryotic Argonaute proteins. My research provided, for the first time, structural insights how guide RNA binding is achieved by eukaryotic Argonaute proteins. In particular, I showed how the AGO MID domain specifically recognizes the 5′-end of the guide RNA (Boland et al., 2010).

Furthermore, I also determined the crystal structure of the first MID-PIWI lobe of a eukaryotic Argonaute protein (Boland et al., 2011). This structure led to the identification of several eukaryote-specific loops present in the PIWI domain (i.e. the catalytic domain) which are important for guide RNA binding. I showed that guide RNA binding is achieved by a composite binding pocket that forms at the MID-PIWI interface and that the integrity of this pocket is crucial for guide RNA binding (Boland et al., 2011). These results challenged a previously published model on allosteric control of eukaryotic AGOs and explained the long-range effects on guide RNA binding observed upon mutation of a critical aspartate residue in the MID-PIWI interface.

The crystal structure of the human AGO MID domain (Frank et al., 2010), published at around the same time as the *Nc* QDE-2 MID domain (Boland et al., 2010), confirmed the results and identified a so-called nucleotide specificity loop, which explains the observed bias for uridine at the 5′-position of the guide RNA.

Lately, several crystal structures of the full-length human AGO1 and AGO2 proteins were published that confirmed and extended our conclusions drawn from the *Neurospora crassa* QDE-2 MID-PIWI lobe (Elkayam et al., 2012; Faehnle et al., 2013; Kuhn and Joshua-Tor, 2013; Nakanishi et al., 2013; Schirle and MacRae, 2012). The crystal structure of human AGO1 revealed that loop L2 is indeed, like predicted in our study,

important for precisely positioning the RNA duplex onto the catalytic site and hence for the slicer activity of the protein.

Most importantly, Schirle & MacRae identified two possible GW interaction sites located in the PIWI domain by co-crystallizing free L-tryptophan, thereby uncovering molecular details of the interaction between GW182/TNRC6 proteins and human AGO2. Very recent work by Pfaff and co-workers (Pfaff et al., 2013), deepened our understanding in the GW182-Argonaute protein interactions by using a peptide scanning approach as well as NMR studies, to identify motifs present in TNRC6B crucial for Argonaute binding. The study led to the identification of two tryptophan residues present in TNRC6B (W623 and W634), being mainly required for AGO2 binding. It was proposed that selection for the correct tryptophan residues for binding to AGO2 is mediated by nonspecific interactions of the flanking regions, in particular of W623. However, to fully understand the details of this interaction between GW182/TNRC6 proteins and AGO proteins in molecular terms, crystal structures of binary complexes of human AGO2 in complex with GW182 peptides will be indispensable.

Finally, the crystal structure of the *Nc* MID-PIWI lobe revealed the presence of a eukaryote-specific α-helical insertion located right above the RNA binding groove. This C-terminal insertion is disordered in the human AGO2 crystal structures; however, given its location, this insertion might serve to discriminate between a bound single stranded substrate and a guide-target duplex. Furthermore this insertion could provide a binding platform for eukaryote-specific binding partners of AGO proteins. Future structural studies on RISC complexes with different binding partners will be critical to expand our knowledge on the mechanisms of RNAi.

In the second half of my PhD thesis, I was working on the structure and the assembly of the two major cytoplasmic deadenylase complexes, namely the PAN2-PAN3 complex and the CCR4-NOT complex. The structure of the PAN3 adaptor protein

(Christie et al., 2013) uncovered the molecular basis for the described self-dimerization, which is mediated by an unpredicted, central coiled coil domain that links the N-terminal pseudokinase domain (PK) to a C-terminal knob (CK) domain. The CK domain has been identified to mediate binding to the catalytic subunit PAN2 and this knowledge was further deepened in our study, through specific point mutations that abolish PAN2 interaction.

The fact that PAN3 homodimerizes in solution and in the crystal environment and that dimerization is conserved throughout evolution is intriguing and immediately suggests that homodimerization is important for PAN3 function. Indeed dimerization is crucial for binding to GW182/TNRC6 proteins, a protein partner only present in higher eukaryotes. In contrast, PAN2 binding does not require dimerization and the isolated CK domain is sufficient for the interaction (see Christie et al., 2013 Figure S4 **C-F**). This result indicates that one PAN3 homodimer could potentially interact with two PAN2 proteins simultaneously. However, the crystal structures of *Nc* and *Dm* PAN3 also revealed that homodimerization is mediated by an asymmetric coiled coil domain, with only one helix being kinked. As a consequence the two C-terminal CK domains are posed into asymmetric structural environments with likely consequences on PAN2 binding affinity. Therefore it might be well possible that one PAN3 homodimer binds only one PAN2 molecule. Future structural studies on the entire PAN2-PAN3 complex will reveal the stoichiometry of the PAN2-PAN3 complex.

The pseudokinase domain of PAN3 adopts a typical bilobal kinase domain fold, although many structural motifs present in canonical kinases are absent in PAN3, which explains the loss of enzymatic activity (see Christie et al., 2013 Figure S4**A**). Importantly, PAN3 proteins however are still capable of binding nucleotides, in a Mg^{2+}-dependent manner (see Christie et al., 2013 Table S2). This suggests that PAN3 could possibly act as an ATP sensor and, for example, couple deadenylation to the energetic state of the cell by conformational changes that regulate the activity of the PAN2-PAN3 complex.

In collaborations, I was also able to obtain structural insights into the assembly of the CCR4-NOT complex by determining the structures of several sub-complexes. These are the NOT-module core formed by the C-termini of CNOT1, CNOT2 and CNOT3 (Boland et al., 2013), the CNOT1 CN9BD in complex with CNOT9 (Chen, 2014) and the CNOT1 MIF4G domain in complex with DDX6 (Chen, 2014).

The structure of the NOT-module core revealed that intrinsically disordered regions (IDRs) present in CNOT2 and CNOT3 are absolutely essential for mediating the assembly of the NOT-module core, which is formed by the C-termini of CNOT1, CNOT2 and CNOT3. Interestingly these IDRs remain largely devoid of any secondary structure, even upon binding to the respective protein partner.

The structure of the minimal NOT-module is of particular relevance because many RNA-associated proteins (including GW182, NANOS and CUP) interact with the C-terminal regions of CNOT1, CNOT2 and CNOT3, to recruit the CCR4-NOT complex to certain mRNA targets (see chapter 4.2). Future structural studies on the NOT-module components in complex with RNA-binding proteins will unveil the molecular details of these interactions.

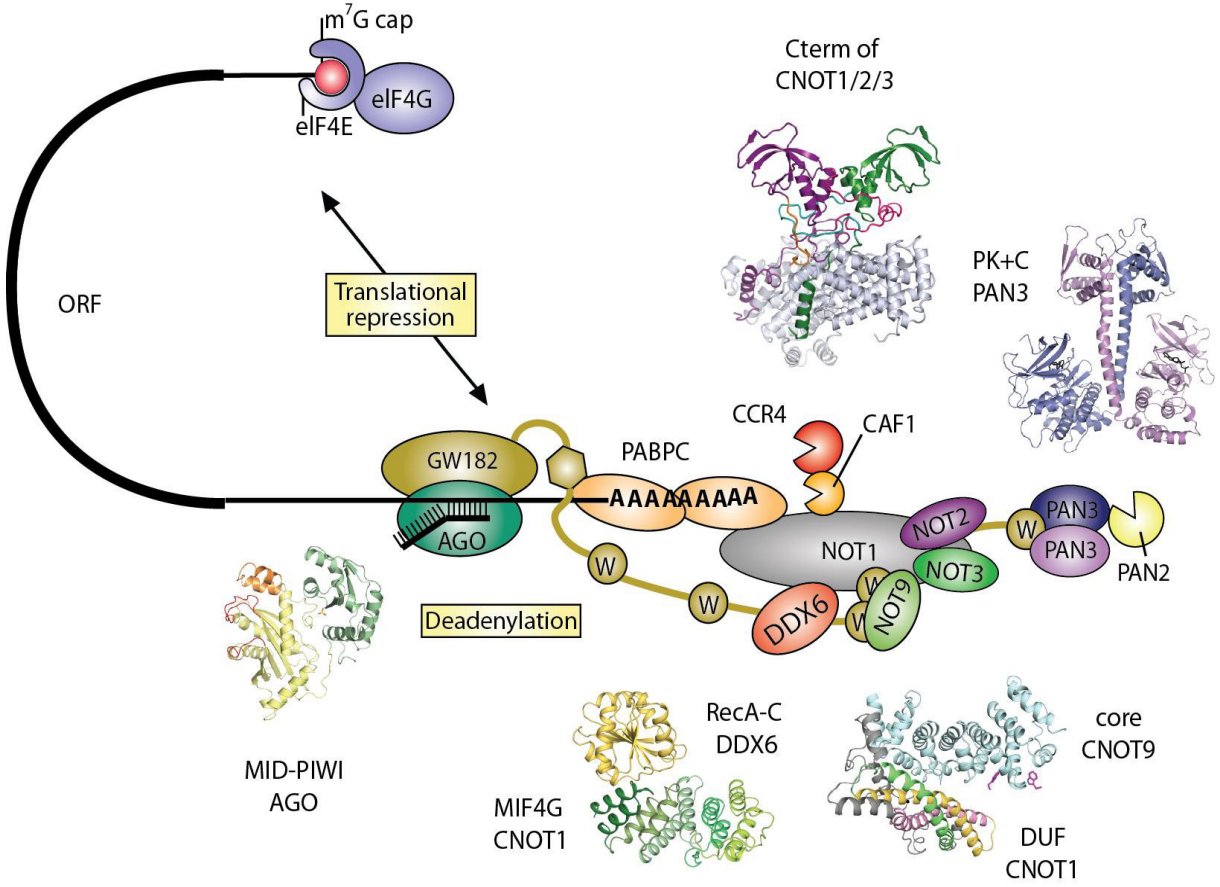
To address the question how the CCR4-NOT complex is recruited to miRNA targets, we solved the structure of a binary complex containing the CNOT1 CN9BD and the CNOT9 ARM domain. Surprisingly, the structure led to the identification of tandem tryptophan binding pockets located in CNOT9. I have validated these tandem W-binding pockets as interaction sites for W-rich proteins *in vitro*. Interestingly the spatial arrangement of the two tryptophan binding pockets is reminiscent of the tryptophan binding pockets identified in the human AGO2 (Schirle and MacRae, 2012). In both proteins the distance between the tandem tryptophan binding pockets is roughly 20-25 Å, a distance that can be spanned by an extended 8-10 amino acid long peptide. This distance is a typical spacer distance between two tryptophan residues in GW182/TNRC6 proteins.

This observation raises the question, why tryptophan residues in the N-terminus of GW182 proteins mainly interact with AGOs, whereas tryptophan residues present in the silencing domain (a C-termianl region of GW182 proteins; see Figure 7) mainly interact with the two cytosolic deadenylases. The potential contribution of residues flanking the tryptophan residue is nicely demonstrated by the structure of *Dm* PAN3 bound to a TWS-containing peptide (stemming from a neighboring symmetry mate) (Christie et al., 2013). The flanking threonine and serine residues stabilize the peptide in a bent conformation, thereby exposing the tryptophan residue. A peptide scanning approach (Pfaff et al., 2013) identified a subset of W-motifs present in TNRC6B and allowed analysis of the flanking residues. It was found that the contact with AGOs is mainly mediated by the tryptophan residues and that small or hydrophilic residues are tolerated as flanking residues whereas bulky side-chains are not. Clearly, as for the AGO-GW182 interaction, future studies co-crystallizing tryptophan rich peptides will provide molecular insights into the specificity of the interaction between AGOs and the deadenylases with GW182/TNRC6 proteins.

Finally, the structure of the CNOT1 MIF4G in complex witht the DDX6 RecA-C domain addressed the question of the repressive action on translation by the CCR4-NOT complex. A recent model by Meijer and co-workers (Meijer et al., 2013) proposed that the described repressive activity of the CCR4-NOT complex on translation is mediated through a direct interaction of the CNOT1 MIF4G domain with eIF4A2, reminiscent to the interaction of eIF4G with eIF4A. It was proposed that the CNOT1 MIF4G domain competes with eIF4G for eIF4A2 binding, however without demonstrating a direct interaction between CNOT1 and eIF4A2. We refute this hypothesis and demonstrate that the MIF4G domain of CNOT1 does not bind eIF4A2 but instead binds another helicase protein, namely the translational repressor DDX6. Additionally, the translational repressor DDX6 has been demonstrated to directly interact with EDC3 and PAT, two components of the decapping machinery as well as with Sde6, a translational

repressor. Therefore our data complete a pathway of direct physical interactions between deadenylation, translational repression and decapping. Future structural studies will reveal how translational repression is achieved mechanistically on an atomic level.

In Figure 19 this work is graphically summarized.



Highlights presented in this thesis:

- The MID domain reveals sRNA 5′-end recognition by eukaryotic AGOs.
- The MID-PIWI interface is crucial for guide RNA binding.
- PAN3 dimerizes via a central coiled coil domain that harbors a W-binding pocket, required for RNA targeting.
- The assembly of the minimal NOT-module is orchestrated by IDRs.
- CNOT9 binds the CNOT1 DUF3819 domain and harbors tandem W-binding pockets, required for RNA targeting.
- DDX6 binds the CNOT1 MIF4G domain, thereby providing a molecular link between decapping and deadenylation.

Figure 19: Structural summary of the work presented in this doctoral thesis. An Argonaute protein recruits through the bridging factor GW182/TNRC6 the two deadenylases PAN2-PAN3 and CCR4-NOT to the mRNA target. The CCR4-NOT complex in turn binds the translational repressor and decapping activator DDX6, thus linking deadenylation and decapping through direct interactions.

6 Author´s contributions to publication

6.1 Crystal structure and ligand binding of the *Nc* QDE-2 MID domain

Boland A, Tritschler F, Heimstädt S, Izaurralde E, Weichenrieder O (2010) Crystal structure and ligand binding of the MID domain of a eukaryotic Argonaute protein. *EMBO Reports* 2010 Jul; 11(7): 522-7.

A. Boland designed the constructs, expressed and purified the *Nc* QDE-2 MID domain as wild-type and selenomethionine substituted versions. He optimized the crystals, collected the diffraction data and contributed to the writing of the manuscript.

6.2 Crystal structure and ligand binding of the *Nc* QDE-2 MID-PIWI domains

Boland A, Huntzinger E, Schmidt S, Izaurralde E, Weichenrieder O (2011) Crystal structure of the MID-PIWI lobe of a eukaryotic Argonaute protein. *PNAS* 2011 Jun; 108(26): 10466-71.

A. Boland designed all constructs including RNA-binding mutants, expressed and purified all native *Nc* QDE-2 MID-PIWI proteins (wild-type and mutant versions) as well as a selenomethionine substituted version. He optimized the crystals, collected the diffraction data, phased the structure experimentally, refined all structures and contributed to the writing of the manuscript. He also performed biochemical assays like guide-RNA binding studies.

6.2 A direct interaction between DCP1 and XRN1 links decapping to degradation

Braun JE, Truffault V, **Boland A**, Huntzinger E, Chang CT, Haas G, Weichenrieder O, Izaurralde E (2012) A direct interaction between DCP1 and XRN1 couples mRNA decapping to 5´ exonucleolytic degradation. *NSMB* 2012 Dec; 19(12): 1324-31.

A. Boland contributed to the purification of the proteins and set up crystallization trials. Furthermore he contributed to the completion of the manuscript including design of mutants, writing of the manuscript and the preparation of structural figures.

6.3 Crystal structure of the PAN3 adaptor protein

Christie M*, **Boland A***, Huntzinger E, Schmidt S, Izaurralde E, Weichenrieder O (2011) Structure of the PAN3 pseudokinase reveals the basis for interactions with the PAN2 deadenylase and the GW182 proteins. Mol Cell. 2013 Aug 8;51(3):360-73.

A. Boland initiated the project, designed all *Nc* PAN3 constructs including a mutated version, expressed and prepared native proteins (different wild-type fragments and the mutant) and selenomethionine substituted protein. He characterized the protein biochemically by using thermal shift assays to demonstrate nucleotide binding and performed the ATPase assay. He obtained and optimized the first PAN3 crystals, collected diffraction data sets, phased the structure experimentally, prepared figures and contributed to the writing of the manuscript.

6.4 Crystal structure of the human NOT-module core and its individual domains

Boland A*, Chen Y*, Raisch T*, Jonas S*, Kuzuoğlu-Öztürk D, Wohlbold L, Weichenrieder O, Izaurralde E (2013) Structure and assembly of the NOT-module of the human CCR4-NOT complex. *NSMB* 2013 Nov; 20(11): 1289-97.

A. Boland established the project in the lab. He designed all constructs used for crystallization including the isolated domains of *Ct* NOT1, *Hs* CNOT2 and *Hs* CNOT3. He also designed the constructs used for the expression of a ternary complex (excluding the *Hs* CNOT1 construct). He contributed to the purification, crystallization and structure determination (data collection, experimental phasing and refinement) of all isolated domains.

Furthermore, he expressed and optimized the purification procedure for the ternary complex and identified a minimal version of the *Hs* CNOT1 C-term. He contributed to the interpretation of data and to the writing of the manuscript.

6.5 Crystal structures of CNOT1-DDX6 and CNOT1-CNOT9 complexes

Chen Y*, **Boland A***, Kuzuoğlu-Öztürk D, Bawankar P, Chang CT, Loh B, Weichenrieder O, Izaurralde E (2013) A DDX6-CNOT1 complex and W-binding pockets in CNOT9 reveal direct links between miRNA target recognition and silencing. Mol Cell. 2014 Jun 5;54(5):737-50.

A. Boland initiated the project and designed several *Hs* CNOT1 DUF3819 constructs. He established the expression and purification of the binary CNOT1-CNOT9 complex and purified several complexes. He contributed to data collection and phased the structure using MR. He further designed binding mutants of CNOT1 that abrogated CNOT9 binding and mutants of CNOT9 that abolished CNOT1 or TNRC6 binding. He performed *in vitro* pull-down assays validating the importance of the indentified W-binding pockets with purified proteins or protein complexes. Finally, he also contributed to the structure determination process of the CNOT1-DDX6 complex including crystal freezing, data collection and structure solution by MR. Again he designed binding mutants, prepared figures and contributed to the writing of the manuscript.

7 Reference List

Ameres, S. L., and Zamore, P. D. (2013). Diversifying microRNA sequence and function. Nat Rev Mol Cell Biol *14*, 475-488.

Astrom, J., Astrom, A., and Virtanen, A. (1991). In vitro deadenylation of mammalian mRNA by a HeLa cell 3' exonuclease. EMBO J *10*, 3067-3071.

Astrom, J., Astrom, A., and Virtanen, A. (1992). Properties of a HeLa cell 3' exonuclease specific for degrading poly(A) tails of mammalian mRNA. J Biol Chem *267*, 18154-18159.

Bai, Y., Salvadore, C., Chiang, Y. C., Collart, M. A., Liu, H. Y., and Denis, C. L. (1999). The CCR4 and CAF1 proteins of the CCR4-NOT complex are physically and functionally separated from NOT2, NOT4, and NOT5. Mol Cell Biol *19*, 6642-6651.

Bartlam, M., and Yamamoto, T. (2010). The structural basis for deadenylation by the CCR4-NOT complex. Protein Cell *1*, 443-452.

Basquin, J., Roudko, V. V., Rode, M., Basquin, C., Seraphin, B., and Conti, E. (2012). Architecture of the nuclease module of the yeast Ccr4-not complex: the Not1-Caf1-Ccr4 interaction. Mol Cell *48*, 207-218.

Bawankar, P., Loh, B., Wohlbold, L., Schmidt, S., and Izaurralde, E. (2013). NOT10 and C2orf29/NOT11 form a conserved module of the CCR4-NOT complex that docks onto the NOT1 N-terminal domain. RNA Biol *10*, 228-244.

Bazzini, A. A., Lee, M. T., and Giraldez, A. J. (2012). Ribosome profiling shows that miR-430 reduces translation before causing mRNA decay in zebrafish. Science *336*, 233-237.

Bernstein, P., Peltz, S. W., and Ross, J. (1989). The poly(A)-poly(A)-binding protein complex is a major determinant of mRNA stability in vitro. Mol Cell Biol *9*, 659-670.

Bernstein, P., and Ross, J. (1989). Poly(A), poly(A) binding protein and the regulation of mRNA stability. Trends Biochem Sci *14*, 373-377.

Bethune, J., Artus-Revel, C. G., and Filipowicz, W. (2012). Kinetic analysis reveals successive steps leading to miRNA-mediated silencing in mammalian cells. EMBO Rep *13*, 716-723.

Boeck, R., Tarun, S., Jr., Rieger, M., Deardorff, J. A., Muller-Auer, S., and Sachs, A. B. (1996). The yeast Pan2 protein is required for poly(A)-binding protein-stimulated poly(A)-nuclease activity. J Biol Chem *271*, 432-438.

Bohmert, K., Camus, I., Bellini, C., Bouchez, D., Caboche, M., and Benning, C. (1998). AGO1 defines a novel locus of Arabidopsis controlling leaf development. EMBO J *17*, 170-180.

Boland, A., Chen, Y., Raisch, T., Jonas, S., Kuzuoglu-Ozturk, D., Wohlbold, L., Weichenrieder, O., and Izaurralde, E. (2013). Structure and assembly of the NOT module of the human CCR4-NOT complex. Nat Struct Mol Biol *20*, 1289-1297.

Boland, A., Huntzinger, E., Schmidt, S., Izaurralde, E., and Weichenrieder, O. (2011). Crystal structure of the MID-PIWI lobe of a eukaryotic Argonaute protein. Proc Natl Acad Sci U S A *108*, 10466-10471.

Boland, A., Tritschler, F., Heimstadt, S., Izaurralde, E., and Weichenrieder, O. (2010). Crystal structure and ligand binding of the MID domain of a eukaryotic Argonaute protein. EMBO Rep *11*, 522-527.

Braun, J. E., Huntzinger, E., Fauser, M., and Izaurralde, E. (2011). GW182 proteins directly recruit cytoplasmic deadenylase complexes to miRNA targets. Mol Cell *44*, 120-133.

Braun, J. E., Huntzinger, E., and Izaurralde, E. (2013). The role of GW182 proteins in miRNA-mediated gene silencing. Adv Exp Med Biol *768*, 147-163.

Braun, J. E., Truffault, V., Boland, A., Huntzinger, E., Chang, C. T., Haas, G., Weichenrieder, O., Coles, M., and Izaurralde, E. (2012). A direct interaction between DCP1 and XRN1 couples mRNA decapping to 5' exonucleolytic degradation. Nat Struct Mol Biol *19*, 1324-1331.

Bryant, T. N., Watson, H. C., and Wendell, P. L. (1974). Structure of yeast phosphoglycerate kinase. Nature *247*, 14-17.

Buchan, D. W., Minneci, F., Nugent, T. C., Bryson, K., and Jones, D. T. (2013). Scalable web services for the PSIPRED Protein Analysis Workbench. Nucleic Acids Res *41*, W349-357.

Cade, R. M., and Errede, B. (1994). MOT2 encodes a negative regulator of gene expression that affects basal expression of pheromone-responsive genes in Saccharomyces cerevisiae. Mol Cell Biol *14*, 3139-3149.

Cali, B. M., Kuchma, S. L., Latham, J., and Anderson, P. (1999). smg-7 is required for mRNA surveillance in Caenorhabditis elegans. Genetics *151*, 605-616.

Chekulaeva, M., Mathys, H., Zipprich, J. T., Attig, J., Colic, M., Parker, R., and Filipowicz, W. (2011). miRNA repression involves GW182-mediated recruitment of CCR4-NOT through conserved W-containing motifs. Nat Struct Mol Biol *18*, 1218-1226.

Chen, C. Y., Gherzi, R., Ong, S. E., Chan, E. L., Raijmakers, R., Pruijn, G. J., Stoecklin, G., Moroni, C., Mann, M., and Karin, M. (2001). AU binding proteins recruit the exosome to degrade ARE-containing mRNAs. Cell *107*, 451-464.

Chen, C. Y., and Shyu, A. B. (2011). Mechanisms of deadenylation-dependent decay. Wiley Interdiscip Rev RNA *2*, 167-183.

Chen, Y., Boland, A., Kuzuoğlu-Öztürk, D., Loh, B., Bawankar, P., Chang, CT., Weichenrieder, O., Izaurralde, E. (2014). Tandem W-binding pockets in CNOT9 and a DDX6-CNOT1 complex complete a pathway of direct molecular interactions that links miRNA target recognition to mRNA silencing

Chiu, Y. L., and Rana, T. M. (2002). RNAi in human cells: basic structural and functional features of small interfering RNA. Mol Cell *10*, 549-561.

Christie, M., Boland, A., Huntzinger, E., Weichenrieder, O., and Izaurralde, E. (2013). Structure of the PAN3 pseudokinase reveals the basis for interactions with the PAN2 deadenylase and the GW182 proteins. Mol Cell *51*, 360-373.

Chu, C. Y., and Rana, T. M. (2006). Translation repression in human cells by microRNA-induced gene silencing requires RCK/p54. PLoS Biol *4*, e210.

Collart, M. A., and Panasenko, O. O. (2012). The Ccr4--not complex. Gene *492*, 42-53.

Collart, M. A., and Struhl, K. (1993). CDC39, an essential nuclear protein that negatively regulates transcription and differentially affects the constitutive and inducible HIS3 promoters. EMBO J *12*, 177-186.

Collart, M. A., and Struhl, K. (1994). NOT1(CDC39), NOT2(CDC36), NOT3, and NOT4 encode a global-negative regulator of transcription that differentially affects TATA-element utilization. Genes Dev *8*, 525-537.

Conti, E., and Izaurralde, E. (2005). Nonsense-mediated mRNA decay: molecular insights and mechanistic variations across species. Curr Opin Cell Biol *17*, 316-325.

Decker, C. J., and Parker, R. (1993). A turnover pathway for both stable and unstable mRNAs in yeast: evidence for a requirement for deadenylation. Genes Dev *7*, 1632-1643.

Diebold, M. L., Fribourg, S., Koch, M., Metzger, T., and Romier, C. (2011). Deciphering correct strategies for multiprotein complex assembly by co-expression: application to complexes as large as the histone octamer. J Struct Biol *175*, 178-188.

Ding, L., and Han, M. (2007). GW182 family proteins are crucial for microRNA-mediated gene silencing. Trends Cell Biol *17*, 411-416.

Ding, L., Spencer, A., Morita, K., and Han, M. (2005). The developmental timing regulator AIN-1 interacts with miRISCs and may target the argonaute protein ALG-1 to cytoplasmic P bodies in C. elegans. Mol Cell *19*, 437-447.

Djuranovic, S., Nahvi, A., and Green, R. (2012). miRNA-mediated gene silencing by translational repression followed by mRNA deadenylation and decay. Science *336*, 237-240.

Djuranovic, S., Zinchenko, M. K., Hur, J. K., Nahvi, A., Brunelle, J. L., Rogers, E. J., and Green, R. (2010). Allosteric regulation of Argonaute proteins by miRNAs. Nat Struct Mol Biol *17*, 144-150.

Doidge, R., Mittal, S., Aslam, A., and Winkler, G. S. (2012). Deadenylation of cytoplasmic mRNA by the mammalian Ccr4-Not complex. Biochem Soc Trans *40*, 896-901.

Eberle, A. B., Lykke-Andersen, S., Muhlemann, O., and Jensen, T. H. (2009). SMG6 promotes endonucleolytic cleavage of nonsense mRNA in human cells. Nat Struct Mol Biol *16*, 49-55.

Eckmann, C. R., Rammelt, C., and Wahle, E. (2011). Control of poly(A) tail length. Wiley Interdiscip Rev RNA *2*, 348-361.

El-Shami, M., Pontier, D., Lahmy, S., Braun, L., Picart, C., Vega, D., Hakimi, M. A., Jacobsen, S. E., Cooke, R., and Lagrange, T. (2007). Reiterated WG/GW motifs form functionally and evolutionarily conserved ARGONAUTE-binding platforms in RNAi-related components. Genes Dev *21*, 2539-2544.

Elbashir, S. M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K., and Tuschl, T. (2001). Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. Nature *411*, 494-498.

Elkayam, E., Kuhn, C. D., Tocilj, A., Haase, A. D., Greene, E. M., Hannon, G. J., and Joshua-Tor, L. (2012). The structure of human argonaute-2 in complex with miR-20a. Cell *150*, 100-110.

Eulalio, A., Tritschler, F., and Izaurralde, E. (2009). The GW182 protein family in animal cells: new insights into domains required for miRNA-mediated gene silencing. RNA *15*, 1433-1442.

Eystathioy, T., Chan, E. K., Tenenbaum, S. A., Keene, J. D., Griffith, K., and Fritzler, M. J. (2002). A phosphorylated cytoplasmic autoantigen, GW182, associates with a unique population of human mRNAs within novel cytoplasmic speckles. Mol Biol Cell *13*, 1338-1351.

Fabian, M. R., Cieplak, M. K., Frank, F., Morita, M., Green, J., Srikumar, T., Nagar, B., Yamamoto, T., Raught, B., Duchaine, T. F., and Sonenberg, N. (2011). miRNA-mediated deadenylation is orchestrated by GW182 through two conserved motifs that interact with CCR4-NOT. Nat Struct Mol Biol *18*, 1211-1217.

Fabian, M. R., Frank, F., Rouya, C., Siddiqui, N., Lai, W. S., Karetnikov, A., Blackshear, P. J., Nagar, B., and Sonenberg, N. (2013). Structural basis for the recruitment of the human CCR4-NOT deadenylase complex by tristetraprolin. Nat Struct Mol Biol *20*, 735-739.

Fabian, M. R., and Sonenberg, N. (2012). The mechanics of miRNA-mediated gene silencing: a look under the hood of miRISC. Nat Struct Mol Biol *19*, 586-593.

Faehnle, C. R., Elkayam, E., Haase, A. D., Hannon, G. J., and Joshua-Tor, L. (2013). The making of a slicer: activation of human Argonaute-1. Cell Rep *3*, 1901-1909.

Faehnle, C. R., and Joshua-Tor, L. (2010). Argonaute MID domain takes centre stage. EMBO Rep *11*, 564-565.

Filipowicz, W. (1978). Functions of the 5,-terminal m7G cap in eukaryotic mRNA. FEBS Lett *96*, 1-11.

Filipowicz, W., Bhattacharyya, S. N., and Sonenberg, N. (2008). Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? Nat Rev Genet *9*, 102-114.

Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., and Mello, C. C. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature *391*, 806-811.

Folta-Stogniew, E., and Williams, K. R. (1999). Determination of molecular masses of proteins in solution: Implementation of an HPLC size exclusion chromatography and laser light scattering service in a core laboratory. J Biomol Tech *10*, 51-63.

Frank, F., Sonenberg, N., and Nagar, B. (2010). Structural basis for 5'-nucleotide base-specific recognition of guide RNA by human AGO2. Nature *465*, 818-822.

Friend, K., Campbell, Z. T., Cooke, A., Kroll-Conner, P., Wickens, M. P., and Kimble, J. (2012). A conserved PUF-Ago-eEF1A complex attenuates translation elongation. Nat Struct Mol Biol *19*, 176-183.

Furuichi, Y., LaFiandra, A., and Shatkin, A. J. (1977). 5'-Terminal structure and mRNA stability. Nature *266*, 235-239.

Gallie, D. R. (1991). The cap and poly(A) tail function synergistically to regulate mRNA translational efficiency. Genes Dev *5*, 2108-2116.

Garces, R. G., Gillon, W., and Pai, E. F. (2007). Atomic model of human Rcd-1 reveals an armadillo-like-repeat protein with in vitro nucleic acid binding properties. Protein Sci *16*, 176-188.

Gatfield, D., and Izaurralde, E. (2004). Nonsense-mediated messenger RNA decay is initiated by endonucleolytic cleavage in Drosophila. Nature *429*, 575-578.

Gavin, A. C., Bosche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J. M., Michon, A. M., Cruciat, C. M., *et al.* (2002). Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature *415*, 141-147.

Ghildiyal, M., Seitz, H., Horwich, M. D., Li, C., Du, T., Lee, S., Xu, J., Kittler, E. L., Zapp, M. L., Weng, Z., and Zamore, P. D. (2008). Endogenous siRNAs derived from transposons and mRNAs in Drosophila somatic cells. Science *320*, 1077-1081.

Ghildiyal, M., Xu, J., Seitz, H., Weng, Z., and Zamore, P. D. (2010). Sorting of Drosophila small silencing RNAs partitions microRNA* strands into the RNA interference pathway. RNA *16*, 43-56.

Grivna, S. T., Beyret, E., Wang, Z., and Lin, H. (2006). A novel class of small RNAs in mouse spermatogenic cells. Genes Dev *20*, 1709-1714.

Guo, L., and Lu, Z. (2010). The fate of miRNA* strand through evolutionary analysis: implication for degradation as merely carrier strand or potential regulatory molecule? PLoS One *5*, e11387.

Hamilton, A. J., and Baulcombe, D. C. (1999). A species of small antisense RNA in posttranscriptional gene silencing in plants. Science *286*, 950-952.

Hanks, S. K., Quinn, A. M., and Hunter, T. (1988). The protein kinase family: conserved features and deduced phylogeny of the catalytic domains. Science *241*, 42-52.

He, X. J., Hsu, Y. F., Zhu, S., Wierzbicki, A. T., Pontes, O., Pikaard, C. S., Liu, H. L., Wang, C. S., Jin, H., and Zhu, J. K. (2009). An effector of RNA-directed DNA methylation in arabidopsis is an ARGONAUTE 4- and RNA-binding protein. Cell *137*, 498-508.

Hodgkin, J., Papp, A., Pulak, R., Ambros, V., and Anderson, P. (1989). A new kind of informational suppression in the nematode Caenorhabditis elegans. Genetics *123*, 301-313.

Hu, H. Y., Yan, Z., Xu, Y., Hu, H., Menzel, C., Zhou, Y. H., Chen, W., and Khaitovich, P. (2009). Sequence features associated with microRNA strand selection in humans and flies. BMC Genomics *10*, 413.

Huntzinger, E., and Izaurralde, E. (2011). Gene silencing by microRNAs: contributions of translational repression and mRNA decay. Nat Rev Genet *12*, 99-110.

Huntzinger, E., Kashima, I., Fauser, M., Sauliere, J., and Izaurralde, E. (2008). SMG6 is the catalytic endonuclease that cleaves mRNAs containing nonsense codons in metazoan. RNA *14*, 2609-2617.

Huntzinger, E., Kuzuoglu-Ozturk, D., Braun, J. E., Eulalio, A., Wohlbold, L., and Izaurralde, E. (2013). The interactions of GW182 proteins with PABP and deadenylases are required for both translational repression and degradation of miRNA targets. Nucleic Acids Res *41*, 978-994.

Irvine, D. V., Zaratiegui, M., Tolia, N. H., Goto, D. B., Chitwood, D. H., Vaughn, M. W., Joshua-Tor, L., and Martienssen, R. A. (2006). Argonaute slicing is required for heterochromatic silencing and spreading. Science *313*, 1134-1137.

Jakymiw, A., Lian, S., Eystathioy, T., Li, S., Satoh, M., Hamel, J. C., Fritzler, M. J., and Chan, E. K. (2005). Disruption of GW bodies impairs mammalian RNA interference. Nat Cell Biol *7*, 1267-1274.

Kinch, L. N., and Grishin, N. V. (2009). The human Ago2 MC region does not contain an eIF4E-like mRNA cap binding motif. Biol Direct *4*, 2.

Kiriakidou, M., Tan, G. S., Lamprinaki, S., De Planell-Saguer, M., Nelson, P. T., and Mourelatos, Z. (2007). An mRNA m7G cap binding-like motif within human Ago2 represses translation. Cell *129*, 1141-1151.

Korner, C. G., and Wahle, E. (1997). Poly(A) tail shortening by a mammalian poly(A)-specific 3'-exoribonuclease. J Biol Chem *272*, 10448-10456.

Kuhn, C. D., and Joshua-Tor, L. (2013). Eukaryotic Argonautes come into focus. Trends Biochem Sci *38*, 263-271.

Kwak, P. B., and Tomari, Y. (2012). The N domain of Argonaute drives duplex unwinding during RISC assembly. Nat Struct Mol Biol *19*, 145-151.

Lau, N. C., Kolkman, A., van Schaik, F. M., Mulder, K. W., Pijnappel, W. W., Heck, A. J., and Timmers, H. T. (2009). Human Ccr4-Not complexes contain variable deadenylase subunits. Biochem J *422*, 443-453.

Lau, N. C., Lim, L. P., Weinstein, E. G., and Bartel, D. P. (2001). An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. Science *294*, 858-862.

Lau, N. C., Seto, A. G., Kim, J., Kuramochi-Miyagawa, S., Nakano, T., Bartel, D. P., and Kingston, R. E. (2006). Characterization of the piRNA complex from rat testes. Science *313*, 363-367.

Le, H., Tanguay, R. L., Balasta, M. L., Wei, C. C., Browning, K. S., Metz, A. M., Goss, D. J., and Gallie, D. R. (1997). Translation initiation factors eIF-iso4G and eIF-4B interact with the poly(A)-binding protein and increase its RNA binding activity. J Biol Chem *272*, 16247-16255.

Lingel, A., Simon, B., Izaurralde, E., and Sattler, M. (2003). Structure and nucleic-acid binding of the Drosophila Argonaute 2 PAZ domain. Nature *426*, 465-469.

Liu, H. Y., Badarinarayana, V., Audino, D. C., Rappsilber, J., Mann, M., and Denis, C. L. (1998). The NOT proteins are part of the CCR4 transcriptional complex and affect gene expression both positively and negatively. EMBO J *17*, 1096-1106.

Liu, J., Carmell, M. A., Rivas, F. V., Marsden, C. G., Thomson, J. M., Song, J. J., Hammond, S. M., Joshua-Tor, L., and Hannon, G. J. (2004). Argonaute2 is the catalytic engine of mammalian RNAi. Science *305*, 1437-1441.

Liu, J., Rivas, F. V., Wohlschlegel, J., Yates, J. R., 3rd, Parker, R., and Hannon, G. J. (2005). A role for the P-body component GW182 in microRNA function. Nat Cell Biol *7*, 1261-1266.

Lowell, J. E., Rudner, D. Z., and Sachs, A. B. (1992). 3'-UTR-dependent deadenylation by the yeast poly(A) nuclease. Genes Dev *6*, 2088-2099.

Lucet, I. S., Babon, J. J., and Murphy, J. M. (2013). Techniques to examine nucleotide binding by pseudokinases. Biochem Soc Trans *41*, 975-980.

Lykke-Andersen, J., and Wagner, E. (2005). Recruitment and activation of mRNA decay enzymes by two ARE-mediated decay activation domains in the proteins TTP and BRF-1. Genes Dev *19*, 351-361.

Ma, J. B., Ye, K., and Patel, D. J. (2004). Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain. Nature *429*, 318-322.

Ma, J. B., Yuan, Y. R., Meister, G., Pei, Y., Tuschl, T., and Patel, D. J. (2005). Structural basis for 5'-end-specific recognition of guide RNA by the A. fulgidus Piwi protein. Nature *434*, 666-670.

Maillet, L., and Collart, M. A. (2002). Interaction between Not1p, a component of the Ccr4-not complex, a global regulator of transcription, and Dhh1p, a putative RNA helicase. J Biol Chem *277*, 2835-2842.

Makarova, K. S., Wolf, Y. I., van der Oost, J., and Koonin, E. V. (2009). Prokaryotic homologs of Argonaute proteins are predicted to function as key components of a novel system of defense against mobile genetic elements. Biol Direct *4*, 29.

Mangus, D. A., Evans, M. C., Agrin, N. S., Smith, M., Gongidi, P., and Jacobson, A. (2004). Positive and negative regulation of poly(A) nuclease. Mol Cell Biol *24*, 5521-5533.

Marcotrigiano, J., Gingras, A. C., Sonenberg, N., and Burley, S. K. (1997). Cocrystal structure of the messenger RNA 5' cap-binding protein (eIF4E) bound to 7-methyl-GDP. Cell *89*, 951-961.

Marzluff, W. F., Wagner, E. J., and Duronio, R. J. (2008). Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. Nat Rev Genet *9*, 843-854.

Mauxion, F., Preve, B., and Seraphin, B. (2013). C2ORF29/CNOT11 and CNOT10 form a new module of the CCR4-NOT complex. RNA Biol *10*, 267-276.

Meijer, H. A., Kong, Y. W., Lu, W. T., Wilczynska, A., Spriggs, R. V., Robinson, S. W., Godfrey, J. D., Willis, A. E., and Bushell, M. (2013). Translational repression and eIF4A2 activity are critical for microRNA-mediated gene regulation. Science *340*, 82-85.

Miller, J. E., and Reese, J. C. (2012). Ccr4-Not complex: the control freak of eukaryotic cells. Crit Rev Biochem Mol Biol *47*, 315-333.

Ming, D., Wall, M. E., and Sanbonmatsu, K. Y. (2007). Domain motions of Argonaute, the catalytic engine of RNA interference. BMC Bioinformatics *8*, 470.

Morel, A. P., Sentis, S., Bianchin, C., Le Romancer, M., Jonard, L., Rostan, M. C., Rimokh, R., and Corbo, L. (2003). BTG2 antiproliferative protein interacts with the human CCR4 complex existing in vivo in three cell-cycle-regulated forms. J Cell Sci *116*, 2929-2936.

Moser, M. J., Holley, W. R., Chatterjee, A., and Mian, I. S. (1997). The proofreading domain of Escherichia coli DNA polymerase I and other DNA and/or RNA exonuclease domains. Nucleic Acids Res *25*, 5110-5118.

Moussian, B., Schoof, H., Haecker, A., Jurgens, G., and Laux, T. (1998). Role of the ZWILLE gene in the regulation of central shoot meristem cell fate during Arabidopsis embryogenesis. EMBO J *17*, 1799-1809.

Muller, M., Weigand, J. E., Weichenrieder, O., and Suess, B. (2006). Thermodynamic characterization of an engineered tetracycline-binding riboswitch. Nucleic Acids Res *34*, 2607-2617.

Munroe, D., and Jacobson, A. (1990). mRNA poly(A) tail, a 3' enhancer of translational initiation. Mol Cell Biol *10*, 3441-3455.

Nakanishi, K., Ascano, M., Gogakos, T., Ishibe-Murakami, S., Serganov, A. A., Briskin, D., Morozov, P., Tuschl, T., and Patel, D. J. (2013). Eukaryote-specific insertion elements control human ARGONAUTE slicer activity. Cell Rep *3*, 1893-1900.

Napoli, C., Lemieux, C., and Jorgensen, R. (1990). Introduction of a Chimeric Chalcone Synthase Gene into Petunia Results in Reversible Co-Suppression of Homologous Genes in trans. Plant Cell *2*, 279-289.

Neff, A. T., Wilusz, C.J., and Wilusz, J. (2013). mRNA Stability. Meyers: Encyclopedia of Molecular Cell Biology and Molecular Medicine: RNA biology *2nd Edition*.

Olovnikov, I., Chan, K., Sachidanandam, R., Newman, D. K., and Aravin, A. A. (2013). Bacterial argonaute samples the transcriptome to identify foreign DNA. Mol Cell *51*, 594-605.

Parker, J. S., Roe, S. M., and Barford, D. (2004). Crystal structure of a PIWI protein suggests mechanisms for siRNA recognition and slicer activity. EMBO J *23*, 4727-4737.

Parker, J. S., Roe, S. M., and Barford, D. (2005). Structural insights into mRNA recognition from a PIWI domain-siRNA guide complex. Nature *434*, 663-666.

Petit, A. P., Wohlbold, L., Bawankar, P., Huntzinger, E., Schmidt, S., Izaurralde, E., and Weichenrieder, O. (2012). The structural basis for the interaction between the CAF1

nuclease and the NOT1 scaffold of the human CCR4-NOT deadenylase complex. Nucleic Acids Res *40*, 11058-11072.

Pfaff, J., Hennig, J., Herzog, F., Aebersold, R., Sattler, M., Niessing, D., and Meister, G. (2013). Structural features of Argonaute-GW182 protein interactions. Proc Natl Acad Sci U S A *110*, E3770-3779.

Phetrungnapha, A., Panyim, S., and Ongvarrasopone, C. (2013). Penaeus monodon Tudor staphylococcal nuclease preferentially interacts with N-terminal domain of Argonaute-1. Fish Shellfish Immunol *34*, 875-884.

Pratt, A. J., and MacRae, I. J. (2009). The RNA-induced silencing complex: a versatile gene-silencing machine. J Biol Chem *284*, 17897-17901.

Proweller, A., and Butler, J. S. (1997). Ribosome concentration contributes to discrimination against poly(A)- mRNA during translation initiation in Saccharomyces cerevisiae. J Biol Chem *272*, 6004-6010.

Quesada, V., Diaz-Perales, A., Gutierrez-Fernandez, A., Garabaya, C., Cal, S., and Lopez-Otin, C. (2004). Cloning and enzymatic analysis of 22 novel human ubiquitin-specific proteases. Biochem Biophys Res Commun *314*, 54-62.

Rehwinkel, J., Behm-Ansmant, I., Gatfield, D., and Izaurralde, E. (2005). A crucial role for GW182 and the DCP1:DCP2 decapping complex in miRNA-mediated gene silencing. RNA *11*, 1640-1647.

Rice, L. M., Earnest, T. N., and Brunger, A. T. (2000). Single-wavelength anomalous diffraction phasing revisited. Acta Crystallogr D Biol Crystallogr *56*, 1413-1420.

Sachs, A. B., and Deardorff, J. A. (1992). Translation initiation requires the PAB-dependent poly(A) ribonuclease in yeast. Cell *70*, 961-973.

Sandler, H., Kreth, J., Timmers, H. T., and Stoecklin, G. (2011). Not1 mediates recruitment of the deadenylase Caf1 to mRNAs targeted for degradation by tristetraprolin. Nucleic Acids Res *39*, 4373-4386.

Sandler, H., and Stoecklin, G. (2008). Control of mRNA decay by phosphorylation of tristetraprolin. Biochem Soc Trans *36*, 491-496.

Schirle, N. T., and MacRae, I. J. (2012). The crystal structure of human Argonaute2. Science *336*, 1037-1040.

Schweingruber, C., Rufener, S. C., Zund, D., Yamashita, A., and Muhlemann, O. (2013). Nonsense-mediated mRNA decay - mechanisms of substrate mRNA recognition and degradation in mammalian cells. Biochim Biophys Acta *1829*, 612-623.

Shabalina, S. A., and Koonin, E. V. (2008). Origins and evolution of eukaryotic RNA interference. Trends Ecol Evol *23*, 578-587.

Sharova, L. V., Sharov, A. A., Nedorezov, T., Piao, Y., Shaik, N., and Ko, M. S. (2009). Database for mRNA half-life of 19 977 genes obtained by DNA microarray analysis of pluripotent and differentiating mouse embryonic stem cells. DNA Res *16*, 45-58.

Soding, J., Biegert, A., and Lupas, A. N. (2005). The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res *33*, W244-248.

Song, J. J., Liu, J., Tolia, N. H., Schneiderman, J., Smith, S. K., Martienssen, R. A., Hannon, G. J., and Joshua-Tor, L. (2003). The crystal structure of the Argonaute2 PAZ domain reveals an RNA binding motif in RNAi effector complexes. Nat Struct Biol *10*, 1026-1032.

Song, J. J., Smith, S. K., Hannon, G. J., and Joshua-Tor, L. (2004). Crystal structure of Argonaute and its implications for RISC slicer activity. Science *305*, 1434-1437.

Suzuki, A., Igarashi, K., Aisaki, K., Kanno, J., and Saga, Y. (2010). NANOS2 interacts with the CCR4-NOT deadenylation complex and leads to suppression of specific RNAs. Proc Natl Acad Sci U S A *107*, 3594-3599.

Suzuki, A., Saba, R., Miyoshi, K., Morita, Y., and Saga, Y. (2012). Interaction between NANOS2 and the CCR4-NOT deadenylation complex is essential for male germ cell development in mouse. PLoS One *7*, e33558.

Temme, C., Zhang, L., Kremmer, E., Ihling, C., Chartier, A., Sinz, A., Simonelig, M., and Wahle, E. (2010). Subunits of the Drosophila CCR4-NOT complex and their roles in mRNA deadenylation. RNA *16*, 1356-1370.

Till, S., and Ladurner, A. G. (2007). RNA Pol IV plays catch with Argonaute 4. Cell *131*, 643-645.

Till, S., Lejeune, E., Thermann, R., Bortfeld, M., Hothorn, M., Enderle, D., Heinrich, C., Hentze, M. W., and Ladurner, A. G. (2007). A conserved motif in Argonaute-interacting proteins mediates functional interactions through the Argonaute PIWI domain. Nat Struct Mol Biol *14*, 897-903.

Tolia, N. H., and Joshua-Tor, L. (2007). Slicer and the argonautes. Nat Chem Biol *3*, 36-43.

Tomari, Y., Matranga, C., Haley, B., Martinez, N., and Zamore, P. D. (2004). A protein sensor for siRNA asymmetry. Science *306*, 1377-1380.

Tritschler, F., Braun, J. E., Eulalio, A., Truffault, V., Izaurralde, E., and Weichenrieder, O. (2009). Structural basis for the mutually exclusive anchoring of P body components EDC3 and Tral to the DEAD box protein DDX6/Me31B. Mol Cell *33*, 661-668.

Tucker, M., Valencia-Sanchez, M. A., Staples, R. R., Chen, J., Denis, C. L., and Parker, R. (2001). The transcription factor associated Ccr4 and Caf1 proteins are components of the major cytoplasmic mRNA deadenylase in Saccharomyces cerevisiae. Cell *104*, 377-386.

Uchida, N., Hoshino, S., and Katada, T. (2004). Identification of a human cytoplasmic poly(A) nuclease complex stimulated by poly(A)-binding protein. J Biol Chem *279*, 1383-1391.

Wahle, E. (1991). A novel poly(A)-binding protein acts as a specificity factor in the second phase of messenger RNA polyadenylation. Cell *66*, 759-768.

Wahle, E., and Winkler, G. S. (2013). RNA decay machines: deadenylation by the Ccr4-not and Pan2-Pan3 complexes. Biochim Biophys Acta *1829*, 561-570.

Wang, Y., Juranek, S., Li, H., Sheng, G., Tuschl, T., and Patel, D. J. (2008a). Structure of an argonaute silencing complex with a seed-containing guide DNA and target RNA duplex. Nature *456*, 921-926.

Wang, Y., Juranek, S., Li, H., Sheng, G., Wardle, G. S., Tuschl, T., and Patel, D. J. (2009). Nucleation, propagation and cleavage of target RNAs in Ago silencing complexes. Nature *461*, 754-761.

Wang, Y., Sheng, G., Juranek, S., Tuschl, T., and Patel, D. J. (2008b). Structure of the guide-strand-containing argonaute silencing complex. Nature *456*, 209-213.

Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M., and Barton, G. J. (2009). Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics *25*, 1189-1191.

Wiederhold, K., and Passmore, L. A. (2010). Cytoplasmic deadenylation: regulation of mRNA fate. Biochem Soc Trans *38*, 1531-1536.

Yamashita, A., Chang, T. C., Yamashita, Y., Zhu, W., Zhong, Z., Chen, C. Y., and Shyu, A. B. (2005). Concerted action of poly(A) nucleases and decapping enzyme in mammalian mRNA turnover. Nat Struct Mol Biol *12*, 1054-1063.

Yan, K. S., Yan, S., Farooq, A., Han, A., Zeng, L., and Zhou, M. M. (2003). Structure and conserved RNA binding of the PAZ domain. Nature *426*, 468-474.

Yuan, Y. R., Pei, Y., Ma, J. B., Kuryavyi, V., Zhadina, M., Meister, G., Chen, H. Y., Dauter, Z., Tuschl, T., and Patel, D. J. (2005). Crystal structure of A. aeolicus argonaute, a site-specific DNA-guided endoribonuclease, provides insights into RISC-mediated mRNA cleavage. Mol Cell *19*, 405-419.

Zekri, L., Huntzinger, E., Heimstadt, S., and Izaurralde, E. (2009). The silencing domain of GW182 interacts with PABPC1 to promote translational repression and degradation of microRNA targets and is required for target release. Mol Cell Biol *29*, 6220-6231.

Zwartjes, C. G., Jayne, S., van den Berg, D. L., and Timmers, H. T. (2004). Repression of promoter activity by CNOT2, a subunit of the transcription regulatory Ccr4-not complex. J Biol Chem *279*, 10848-10854.

8 Abbreviations

Å	Ångström (1 Å = 0,1 nm = 10^{-10} m)
β-Me	β-Mercaptoethanol
Af	*Archaeoglobus fulgidus*
AGO	Argonaute
ARE	AU-rich elements
AMP	adenosine monophosphate
AMPPNP	Adenosine 5′-(β,γ-imido)triphosphate
ATP	adenosine triphosphate
ATPγS	Adenosine-5'-(γ-thio)-triphosphate
ASU	molecules per asymmetric unit
CAF	CCR4-associated factor
CCR4	Carbon catabolite repressor protein 4
Ct	*Chaetomium thermophilum*
CRISPR	clustered regularly interspaced short palindromic repeats
C-term	C-terminal
cDNA	complementary DNA
DALI	Distance-matrix ALIgnment
Dm	*Drosophila melanogaster*
DCP1	decapping protein 1
DCP2	decapping protein 2
DDX6	DEAD box polypeptide 6
DEAD box	Asp-Glu-Ala-Asp box
DEDD	Asp-Glu-Asp-Asp
Dhh1	DExD/H-box helicase 1
DNA	deoxyribonucleic acid
DTT	dithiothreitol
DUF	domain of unknown function

E. coli	*Escherichia coli*
EDC3	enhancer of decapping 3
EDC4	enhancer of decapping 4
EDTA	ethylenediaminetetraacetic acid
EEP	endonuclease-exonuclease-phosphatase homology domain
eIF	eukaryotic translation initiation factor
FOM	figure of merit
GST	glutathione-S-transferase
GW-repeat	Gly-Trp-repeat
GW182	Gly-Trp repeat containing protein of 182 kDa size
HEAT	Huntingtin, ef3, protein phosphates 2A, TOR1
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
Hs	*Homo sapiens*
IDR	intrinsically disordered region
LSm	Sm-like
m^7G	7-methyl-guanosine
MAD	multiple-wavelength anomalous dispersion
MALLS	multiangle laser-light scattering
MBP	maltose binding protein
Me31B	maternal expression at 31B
MID	Middle
miRISC	miRNA induced silencing complex
miRNA	microRNA
MR	molecular replacement
mRNA	messenger ribonucleic acid
mRNP	messenger ribonucleoprotein
Mut	mutant
Nc	*Neurospora crassa*
Ni-NTA	nickel-nitrilotriacetate
NMD	nonsense-mediated mRNA decay

N-term	N-terminal
NOT	Negative regulator of transcription homolog
nt	nucleotides
ORF	open reading frame
PABPC	cytoplasmic poly(A) binding protein
PABPN	nuclear poly(A) binding protein
PAM2	PABPC-interacting motif 2
PAN	poly(A) specific ribonuclease subunit homolog
PARN	poly(A) specific ribonuclease
Pat1	protein associated with topoisomerase II
PAZ	PIWI, Argonaute, Zwille/Pinhead
PDB	Protein Data Bank
piRNA	PIWI-interacting RNAs
PIWI	P-element induced wimpy testis
poly(A) tail	poly adenine tail
POP2	Poly(A) ribonuclease POP2
pre-mRNA	precursor mRNA
PTC	premature termination codon
PTGS	post-transcriptional gene silencing
QDE-2	quelling deficient-2
r.m.s.d.	root mean square deviation
R_{cryst} / R_{work}	crystallographic reliability factor
R_{free}	free crystallographic reliability factor
RBP	RNA-binding protein
RCD-1	required for cell differentiation protein 1
RISC	RNA-induced silencing complex
RNA	ribonucleic acid
RNAi	RNA interference
RNP	ribonucleoprotein (particle)
RRM	RNA recognition motif

S-SAD	Sulphur-SAD
Sc	*Saccharomyces cerevisiae*
Sp	*Schizosaccharomyces pombe*
SAD	single-wavelength anomalous dispersion
SD	silencing domain of GW182
SDS-PAGE	sodium dodecylsulfate polyacrylamide gel electrophoresis
siRNA	small interfering RNA
SMG	Suppressor of morphogenesis in genitalia
TEV-protease	tobacco etch virus protease
Tris	tris(hydroxymethyl)aminomethane
TNRC6	Trinucleotide repeat-containing gene 6 protein
Tt	*Thermus thermophilus*
TTP	tristetraprolin
UPF	up-frameshift proteins
UTR	untranslated region
WD	tryptophan-aspartate dipeptide
WD40	domain of ~40 residues terminating with a WD dipeptide
Xt	*Xenopus tropicalis*
XRN1	exoribonuclease 1

9 Appendix

9.1 List of publications

9.1.1 Discussed publications

1) ***Tandem tryptophan-binding pockets in CNOT9 mediate binding to TNRC6 proteins and contribute to miRNA-mediated silencing***

Chen Y*, Boland A*, Kuzuoğlu-Öztürk D*, Bawankar P, Chang CT, Loh B, Weichenrieder O, Izaurralde E.
Mol Cell. 2014 Jun 5;54(5):737-50**, *equal contributions**

2) ***Structure and assembly of the NOT-module of the CCR4-NOT complex***

Boland A*, Chen Y*, Raisch T*, Jonas S*, Kuzuoğlu-Öztürk D, Wohlbold L, Weichenrieder O, Izaurralde E.
Nat Struct Mol Biol. 2013 Nov;20(11):1289-97, ***equal contributions**

3) ***The structure of the dimeric PAN3 pseudokinase reveals the basis for interactions with the PAN2 deadenylase and the GW182/TNRC6 proteins.***

Christie M*, Boland A*, Huntzinger E, Weichenrieder O, Izaurralde E.
Mol Cell. 2013 Aug 8;51(3):360-73**, *equal contributions**

4) ***Crystal structure of the MID-PIWI lobe of a eukaryotic Argonaute protein***

Boland A, Huntzinger E, Schmidt S, Izaurralde E, Weichenrieder O.
Proc Natl Acad Sci U S A. 2011 Jun 28;108(26):10466-71

5) ***Crystal structure and ligand binding of the MID domain of a eukaryotic Argonaute protein***

Boland A, Tritschler F, Heimstädt S, Izaurralde E, Weichenrieder O.
EMBO Rep. 2010 Jul;11(7):522-7

9.1.2 Further publications

6) ***A direct interaction between DCP1 and XRN1 couples mRNA decapping to 5'exonucleolytic degradation***

Braun JE, Truffault V, Boland A, Huntzinger E, Chang CT, Haas G, Weichenrieder O, Coles M, Izaurralde E.
Nat Struct Mol Biol. 2012 Dec;19(12):1324-31

9.1.3 Table of solved crystal structures

List of crystal structures determined independently or in collaboration during this thesis

Crystallized protein	Method	Space group	Resolution in Å	R_{work}/R_{free} in %	PDB entry
Nc MID	SAD	C2 2 2_1	2.20	19.4 / 22.3	2XDY
Nc MID-PIWI	SAD	$P6_3$ 2 2	3.65	23.4 / 25.2	2YHA
Nc MID-PIWI ΔL	MR	$P3_2$ 2 1	1.85	19.7 / 23.6	2YHB
Nc PAN3	SAD	$P6_5$	3.30	22.2 / 26.9	4BWP
Nc PAN3 mutant	MR	$P6_5$	2.85	21.3 / 24.2	4BWX
Hs CNOT3	SAD	$P2_1$ 2_1 2_1	2.40	22.3 / 25.7	4C0G
Hs CNOT2	SAD	$P6_5$ 2 2	2.40	21.9 / 26.3	4C0F
Ct NOT1	SAD	$P2_1$	3.20	21.9 / 25.9	4C0E
Hs CNOT1/2/3	MR	$P2_1$ 2_1 2	3.20	22.6 / 27.3	4C0D
Hs CNOT1/CNOT9	MR	C1 2 1	1.65	16.0 / 17.5	N/A
Hs CNOT1/CNOT9/W	MR	C1 2 1	2.05	16.8 / 20.5	N/A
Hs CNOT1/DDX6	MR	$P2_1$ 2_1 2_1	1.75	16.0 / 19.8	N/A
Nc QIP	S-SAD	Unpublished data			

9.2 Curriculum vitae

Andreas Boland

PERSONAL INFORMATION

Date of birth:	9th January 1983 in Cologne
Nationality:	German
Address:	Department of Biochemistry, Max Planck Institute for Developmental Biology Spemannstrasse 35, D-72076 Tübingen, Germany
Phone:	+49 7071 601 1365
Email:	andreas.boland@tuebingen.mpg.de

EDUCATION

01/2009 – present	**PhD student** Group of Prof. Dr. Elisa Izaurralde Max Planck Institute for Developmental Biology, Tübingen, Germany *"Structural characterisation of eukaryotic mRNA decay factors involved in post-transcriptional gene regulation"*
12/2008	**Final exams in Biology (Diplom)** University of Jena, Germany Subjects: Genetics, Biochemistry, Microbiology Grade: excellent (1.4)
11/2007-12/2008	**Research for degree thesis (Diplomarbeit)** Group of Prof. Dr. Christoph Englert Leibniz Institute for Age Research (FLI), Jena, Germany "*Characterization of the + and –KTS isoforms of the tumor supressor protein WT1*"

2003-2008	**Studies in Biology**
	University of Jena, Germany
2002-2003	**Civil Service**
	Assisted daily work in the Psychiatry in Jena
1993-2002	**Grammar school**
	Durmersheim (Karlsruhe), Germany High School Diploma (Abitur)

CONFERENCE PARTICIPATIONS

June **2013**, 18th Annual Meeting of the RNA Society, Davos, **Speaker**

April **2013**, Meeting of the DFG network FOR855 on "Cytoplasmic regulation of gene expression", University of Cologne, **Speaker**

June **2012**, FASEB Science Research Conference on "Post-transcriptional control of gene expression: Mechanisms of mRNA decay, Steamboat Springs, **Poster**

May **2012**, 17th Annual Meeting of the RNA Society, Ann Arbor, **Speaker**

June **2011**, 16th Annual Meeting of the RNA Society, Kyoto, **Poster**

March **2011**, Keystone meeting, "Mechanism and Biology of Silencing", Monterey, **Poster**

March **2011**, Meeting of the DFG network FOR855 on "Cytoplasmic regulation of gene expression", MPI Tübingen, **Speaker**

October **2010**, The 4th SIROCCO Annual Meeting, EMBL Heidelberg, **Speaker**

May **2010**, 5th Microsymposium on Small RNAs, IMBA Vienna, **Invited speaker**

9.3 Acknowledgements

First of all, I would particularly like to thank my supervisor Prof. Dr. Elisa Izaurralde for giving me the opportunity to work in her laboratory on this very exciting research topic(s) and for her constant support during the last years. I deeply enjoyed the time in your laboratory!

I also would like to thank especially Dr. Oliver Weichenrieder for his patience and enthusiasm in teaching me crystallography and for his highly valued advices in scientific as well as non-scientific questions.

I am very grateful to Prof. Dr. Thilo Stehle for the supervision of this thesis at the Eberhard Karls Universität Tübingen and interesting scientific discussion during this PhD project.

Also I would like to express my gratitude to Dr. Steffen Schmidt for his immense support regarding bioinformatics questions and keeping the required crystallographic software always up to date.
Further, I thank Dr. Felix Tritschler for the initial help with my Argonaute project and Dr. Jörg Braun for numerous inspiring scientific discussions and advices during my PhD. I really enjoyed the time that we shared together in the lab.
I am grateful to Ying Chen, Dr. Eric Huntzinger, Dr. Mary Tara Christie, Tobias Raisch, Duygu Kuzuoğlu-Öztürk, Belinda Loh, Praveen Bawankar, Chung-Te Chang and Dr. Susi Heimstädt for great collaborations in various projects.
I would like to express my gratitude to all past and present members of Department II at the MPI for the nice and very stimulating working atmosphere. Especially, I would like to thank Valentina Ahl, Regina Büttner, Elena Khazina, Stefan Grüner and Anna

Schneider for the nice time that we shared together inside and outside the lab. Dear Anna, I wish you all the best for your future plans and that you stay always that optimistic and cheerful like you did in the past.

Also, I would like to thank in particular all the technical assistants namely Maria Fauser, Sigrun Helms, Catrin Weiler and Marie Palaj as well as all the cleaning staff for providing such an excellent working environment.

Special thanks to the PubQuiz Team including Julia Kamenz, Stefanie Heinrich, Nadine Schmidt, Katharina Sewart, Ole Herud, Fabio Zanini and Andrey Fadeev for all the entertaining events that we shared during the past years.

I would like to cordially thank my friends that accompanied me through the last years, especially Thomas Kempf and Stefanie Peters, Stephan Kukla and Janina Enderle, Marc Pusch, Tobias Negwer as well as Matthias Langenbacher and Corina Suceveanu.

Lastly, I deeply thank my parents Maria Luise and Wilhelm Boland, my brothers Daniel and Sebastian Boland as well as my girlfriend Maren Schröder for their constant and highly appreciated support of my past, present and future plans.

EMBO reports

EMBO open

scientific report

Crystal structure and ligand binding of the MID domain of a eukaryotic Argonaute protein

Andreas Boland, Felix Tritschler, Susanne Heimstädt, Elisa Izaurralde[+] *& Oliver Weichenrieder*[++]

Department of Biochemistry, Max Planck Institute for Developmental Biology, Tübingen, Germany

Argonaute (AGO) proteins are core components of RNA-induced silencing complexes and have essential roles in RNA-mediated gene silencing. They are characterized by a bilobal architecture, consisting of one lobe containing the amino-terminal and PAZ domains and another containing the MID and PIWI domains. Except for the PAZ domain, structural information on eukaryotic AGO domains is not yet available. In this study, we report the crystal structure of the MID domain of the eukaryotic AGO protein QDE-2 from *Neurospora crassa*. This domain adopts a Rossmann-like fold and recognizes the 5′-terminal nucleotide of a guide RNA in a manner similar to its prokaryotic counterparts. The 5′-nucleotide-binding site shares common residues with a second, adjacent ligand-binding site, suggesting a mechanism for the cooperative binding of ligands to the MID domain of eukaryotic AGOs.

Keywords: GW182; miRNAs; RNAi; silencing; siRNAs

EMBO *reports* (2010) **11**, 522–527. doi:10.1038/embor.2010.81

INTRODUCTION

Proteins of the Argonaute (AGO) family have essential roles in RNA-mediated gene silencing mechanisms throughout the eukaryotic lineage. They associate with small non-coding RNAs to form RNA–protein effector complexes called RNA-induced silencing complexes (Tolia & Joshua-Tor, 2007; Jínek & Doudna, 2009). In RNA-induced silencing complexes, the small RNA functions as a sequence-specific guide that directs AGOs to fully or partially complementary target RNAs through base pairing interactions. The target RNA is then silenced at the transcriptional or post-transcriptional level (Tolia & Joshua-Tor, 2007; Jínek & Doudna, 2009).

Department of Biochemistry, Max Planck Institute for Developmental Biology, Spemannstrasse 35, Tübingen D-72076, Germany

[+]Corresponding author. Tel: +49 7071 601 1350; Fax: +49 7071 601 1353; E-mail: elisa.izaurralde@tuebingen.mpg.de

[++]Corresponding author. Tel: +49 7071 601 1358; Fax: +49 7071 601 1353; E-mail: oliver.weichenrieder@tuebingen.mpg.de

Received 30 March 2010; revised 10 May 2010; accepted 12 May 2010; published online 11 June 2010

Structural and biochemical studies showed that AGOs consist of four domains: an amino-terminal domain; a PIWI/Argonaute/Zwille (PAZ) domain that binds to the 3′-end of guide RNAs; the middle (MID) domain that provides a binding pocket for the 5′-phosphate of guide RNAs; and the P-element-induced whimpy testes (PIWI) domain that adopts an RNaseH fold and has endonucleolytic activity in some but not all AGOs (Parker *et al*, 2004, 2005, 2009; Song *et al*, 2004; Ma *et al*, 2005; Yuan *et al*, 2005; Wang *et al*, 2008a,b, 2009). Thus far, structural information on full-length AGOs is only available for the homologous proteins from archaea and eubacteria (Song *et al*, 2004; Ma *et al*, 2005; Parker *et al*, 2005; Yuan *et al*, 2005; Wang *et al*, 2008a,b, 2009), which can use DNA instead of RNA as guide molecules (Ma *et al*, 2005; Parker *et al*, 2005; Yuan *et al*, 2005; Wang *et al*, 2008a,b, 2009).

The function of AGO, however, has only been studied in eukaryotic systems; and, in the microRNA (miRNA) pathway, the precise mechanism by which AGO proteins mediate translational repression of target messenger RNAs (mRNAs) remains controversial (Eulalio *et al*, 2008a). Understanding such silencing mechanisms requires detailed structural information about the domains of eukaryotic AGOs; however, so far, this level of information is only available for the PAZ domain (Lingel *et al*, 2003; Song *et al*, 2003; Yan *et al*, 2003).

In this study, we focused on the structure of a eukaryotic AGO MID domain, because it is thought to perform essential and distinct functions in the miRNA pathway and because a considerable amount of discussion has focused on the MID domain fold (Kiriakidou *et al*, 2007; Kinch & Grishin, 2009; Djuranovic *et al*, 2010). Furthermore, it has been proposed that certain eukaryotic AGO MID domains sequester the mRNA 5′-cap structure (m^7GpppN) of target mRNAs, thereby repressing translation. Initially, this idea was based on the observation that the human AGO2 binds to m^7GTP Sepharose beads (Kiriakidou *et al*, 2007). Recent data by Djuranovic *et al* (2010) supported this concept, indicating that, in addition to the 5′-phosphate-binding pocket for the guide RNA, *Drosophila melanogaster* AGO1 has a second miRNA-dependent site that can bind to nucleotides such as the 5′-cap.

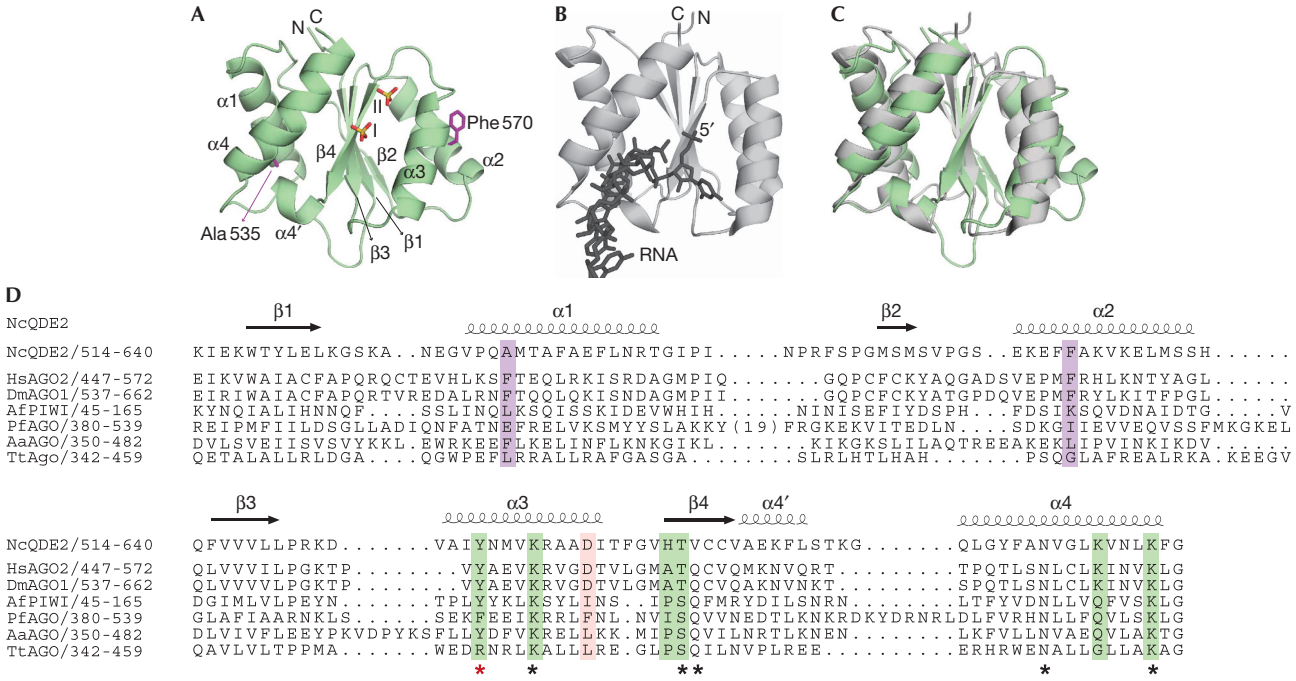


Fig 1 | Structure of the *Neurospora crassa* QDE-2 MID domain. (**A**) Ribbon representation of the MID domain showing the positions of two bound sulphate ions (I and II) as sticks (red: oxygen; yellow: sulphur). Residues corresponding to those proposed by Kiriakidou *et al* (2007) to sandwich the m^{7}GpppN cap are shown as purple sticks. Secondary structure elements are labelled. (**B**) Ribbon representation of an *Af* MID domain in complex with the 5′-end of a guide RNA (sticks), generated from Protein Data Bank entry 2bgg (Parker *et al*, 2005). (**C**) Superposition of the *Nc* QDE-2 MID domain onto the *Af* MID domain, illustrating the conservation of the Rossmann-like fold. (**D**) Structure-based sequence alignment of the *Nc* QDE-2 MID domain with archaeal (*Af*, *Pf*) and eubacterial (*Aa*, *Tt*) homologues, as well as with eukaryotic MID domains involved in the miRNA pathway. Positions involved in coordinating the two sulphate ions are shown in green. Positions proposed by Kiriakidou *et al* (2007) to bind the m^{7}GpppN cap are shown in purple. The position proposed by Djuranovic *et al* (2010) to effect allosteric regulation of miRNA binding is shown in pink. Invariant or highly conserved positions are marked by an asterisk. *Aa*, *Aquifex aeolicus*; *Af*, *Archaeoglobus fulgidus*; MID, middle; miRNA, micro RNA; *Nc*, *Neurospora crassa*; *Pf*, *Pyrococcus furiosus*; QDE-2, quelling deficient 2; *Tt*, *Thermus thermophilus*.

Kiriakidou *et al* (2007) proposed that eukaryotic AGO MID domains fold in a eukaryotic initiation factor 4E (eIF4E)-like manner, sandwiching the m^{7}GpppN cap between two phenylalanine residues (Marcotrigiano *et al*, 1997). By contrast, Djuranovic *et al* (2010) modelled the MID domain onto the Rossmann-like fold of the archaeal and eubacterial homologues. This latter modelling lead the authors to propose allosteric control of cap binding by the presence of the guide RNA 5′-nucleotide in a conserved binding pocket corresponding to that observed in the prokaryotic MID domains.

In this study, we present the crystal structure of the *Neurospora crassa* quelling deficient 2 (QDE-2) MID domain (Fulci & Macino, 2007) as an example of a eukaryotic AGO MID domain. This structure convincingly refutes the idea that eukaryotic MID domains adopt an eIF4E-like fold and shows the guide RNA 5′-phosphate-binding site is conserved. Furthermore, we describe a second, previously unreported ligand-binding site that allows a simple mechanistic explanation for the allosteric effects reported by Djuranovic *et al* (2010).

RESULTS

Structure of a eukaryotic AGO MID domain

The 2.2 Å structure of the *N. crassa* AGO MID domain (*Nc* QDE-2 MID; residues Lys 514–Gly 640) was solved by experimental phasing using a selenomethionine-substituted protein and was refined with excellent stereochemistry (supplementary Table S1 online). The protein adopts a Rossmann-like fold with a central four-stranded parallel β-sheet sandwiched between α-helices in a βαβαβαβαα topology (Fig 1A–D). The structure superimposes well with the structures of previously determined archaeal and eubacterial AGO MID domains, despite the low sequence identity (Fig 1C,D). Among the best-scoring structural relatives is the AGO MID domain from *Archaeoglobus fulgidus* (*Af*; Protein Data Bank (PDB) ID: 1w9h), showing an r.m.s.d. value of 2.6 Å and a sequence identity of 9%. For comparison, the r.m.s.d./sequence identity is 2.2 Å/16% for *Aquifex aeolicus* (*Aa*; PDB ID: 1yvu), 2.7 Å/12% for *Pyrococcus furiosus* (*Pf*; PDB ID: 1u04) and 3.3 Å/14% for *Thermus thermophilus* (*Tt*; PDB ID: 3ho1). By contrast, the sequence identities for eukaryotic homologues are generally greater than 20% (24% for *Homo sapiens* (*Hs*) AGO2, 26% for *D. melanogaster* (*Dm*) AGO1; Fig 1D). The structure of the *Nc* QDE-2 MID domain thus serves as a prototype for eukaryotic AGO MID domains.

The *Nc* QDE-2 MID domain structure has no similarity to eukaryotic eIF4E. Consequently, it is unlikely that eukaryotic AGO MID domains sandwich the m^{7}GpppG mRNA 5′-cap between two aromatic amino-acid side chains in an eIF4E-like

manner. In our structure, the two residues (Ala 535 and Phe 570) corresponding to those proposed to be important for cap binding (Kiriakidou *et al*, 2007) lie more than 25 Å apart (Fig 1A, purple). Earlier sequence analyses and homology modelling reached similar conclusions (Kinch & Grishin, 2009; Djuranovic *et al*, 2010). Furthermore, in previous studies, we showed that double valine substitutions of these residues abolished the silencing activity of *Dm* AGO1 and *Hs* AGO2 by preventing the interaction with both miRNAs and the GW182 protein, which is essential for miRNA-mediated silencing in animal cells (Eulalio *et al*, 2008b).

The 5′-phosphate-binding pocket is highly conserved

Several archaeal AGO proteins have been crystallized in complex with a guide RNA/DNA (Ma *et al*, 2005; Parker *et al*, 2005; Yuan *et al*, 2005; Wang *et al*, 2008a,b, 2009). The respective structures reveal that the 5′-terminal nucleotide of the guide molecule is not available for base pairing with the RNA target but is strongly bent and accommodated in a preformed pocket of the MID domain (Fig 1B). In each case, the base stacks onto an aromatic or arginine side chain (Fig 1D, red asterisk). Most importantly, the 5′-terminal phosphate (a hallmark of small interfering RNAs and miRNAs) is coordinated precisely by several side-chain and main-chain contacts to residues from helix α3, strand β4 and helix α4 (supplementary Fig S1A,B online; Parker *et al*, 2005; Wang *et al*, 2009). Furthermore, in the context of full-length proteins, the carboxyl terminus of the PIWI domain contacts the 5′-phosphate through a metal ion (Parker *et al*, 2005; Wang *et al*, 2009).

When a binary complex containing the *Af* MID and PIWI domains plus a guide RNA (PDB ID: 2bgg; Parker *et al*, 2005) is superposed onto the *Nc* QDE-2 MID domain, the 5′-terminal uridine of the guide RNA is positioned favourably into its positively charged binding pocket, with the base stacking on the aromatic ring of Tyr 595 of the *Nc* QDE-2 MID domain and the 5′-phosphate being coordinated by the Tyr 595 hydroxyl group (Fig 2A,B). Most importantly, the 5′-phosphate superimposes almost perfectly with a sulphate ion (ion I; from the crystallization conditions) that is coordinated additionally by the invariant lysines Lys 599 (helix α3) and Lys 638 (helix α4), as well as by the main-chain nitrogen of Cys 612 (strand β4) and the poorly conserved Lys 634 (helix α4; Fig 2A,B). Conversely, a superposition of the *Nc* QDE-2 MID domain onto the structures of *Af* or *Tt* MID domains places the sulphate ion (I) at the position of the 5′-phosphate of the guide RNA or DNA, respectively; this illustrates that the ligand-binding residues are conserved structurally (supplementary Fig S1A,B online). We conclude that eukaryotic AGO MID domains accommodate the 5′-terminal nucleotide of the guide RNA in a manner similar to their archaeal and eubacterial counterparts.

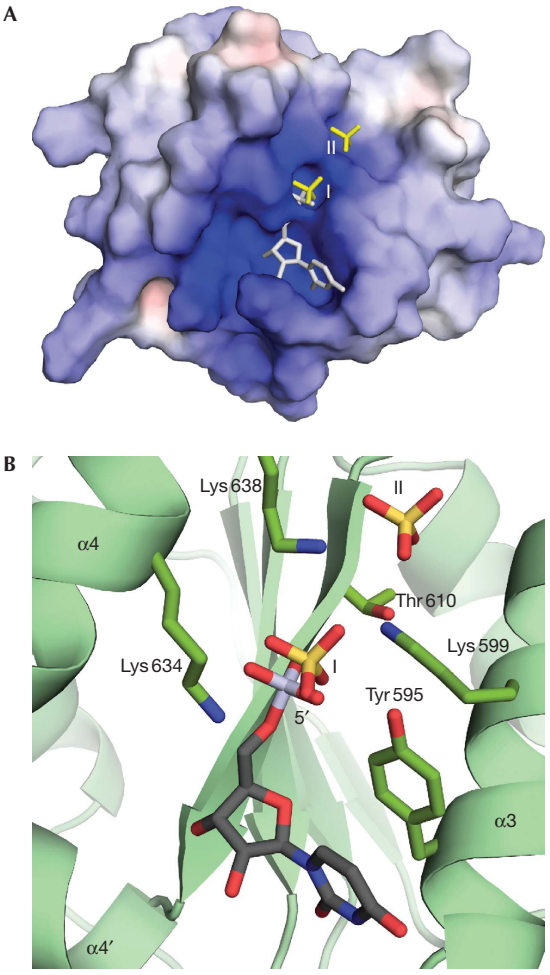


Fig 2 | Conservation of the guide RNA 5′-nucleotide-binding pocket. (**A**) Electrostatic potential mapped onto the molecular surface of the *Nc* QDE-2 MID domain, with the two bound sulphates shown as yellow sticks. Potentials are contoured from −10 (red) to +10 kT/e (blue). The 5′-RNA nucleotide from the superposition with the *Af* MID domain in complex with a guide RNA (Fig 1B,C) is shown as grey sticks. (**B**) Close-up of the RNA 5′-end nucleotide-binding site. The superimposed 5′-uridine from the *Af* complex structure is shown as sticks, together with the two sulphates and important side chains from the *Nc* QDE-2 structure. The uridine base stacks on the aromatic Tyr 595 ring, whereas the 5′-phosphate superimposes with sulphate I. Red: oxygen; blue: nitrogen; yellow: sulphur; violet: phosphorus. *Af*, *Archaeoglobus fulgidus*; MID, middle; *Nc*, *Neurospora crassa*; QDE-2, quelling deficient 2.

A second ligand-binding pocket

The structure of the *Nc* QDE-2 MID domain contains a second sulphate ion (II), only 6.3 Å from the first ion; this second ion is also well coordinated (Fig 3A,B). This second ligand-binding pocket is separated from the first 5′-phosphate-binding pocket by only two invariant lysines, Lys 599 and Lys 638, which thus participate simultaneously in coordinating both sulphate ions (Fig 3A,B). The second sulphate ion is coordinated further by the main-chain and side-chain atoms from Thr 610 (strand β4), which is invariantly a serine or threonine, and by the non-conserved side-chain of His 609 (strand β4; Fig 3A,B). As a consequence, the two sulphate-binding pockets are not independent of each other, but are likely to display positive cooperativity (that is, a ligand binding to either would orient the lysines favourably and promote binding of a ligand to the other site).

The presence of two positively coupled ligand-binding sites on the *Nc* AGO MID domain is particularly interesting in the context of a recent report by Djuranovic *et al* (2010), indicating

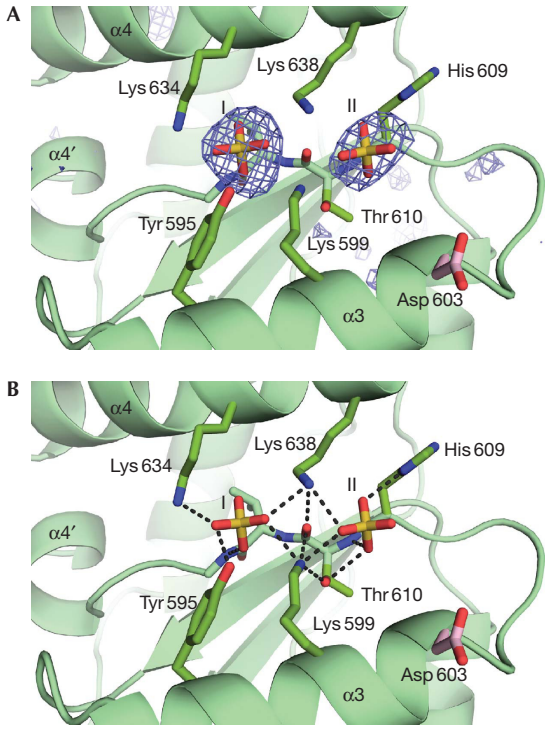


Fig 3 | Coupled coordination of the two sulphate ion ligands. (**A**) Difference density for the sulphate ions. Difference density ($(|F_o|-|F_c|)e(i\alpha_c)$) is contoured at 3.0 sigma values over the mean, where $|F_o|$ is the observed structure factor amplitudes and $|F_c|$ and α_c are structure factor amplitudes and phases calculated from a model prior to including the sulphates. (**B**) Potential hydrogen bonds of the two sulphate ions (I and II) to side-chain and main-chain atoms of the *Nc* QDE-2 MID domain. Relevant side chains and the main chain of strand β4 are shown as sticks. Hydrogen bonds are shown as dotted lines. Asp 603 (pink) corresponds to the aspartate proposed by Djuranovic *et al* (2010) to effect allosteric regulation of miRNA binding. Red: oxygen; blue: nitrogen; yellow: sulphur. MID, middle; miRNA, micro RNA; *Nc*, *Neurospora crassa*; QDE-2, quelling deficient 2.

that miRNA binding to the MID domain of *Dm* AGO1 might be under allosteric control (Djuranovic *et al*, 2010). These authors suggested that the MID domains of certain eukaryotic AGOs (that is, those involved in miRNA-mediated gene regulation) contain a second nucleotide-binding site in addition to the 5′-phosphate-binding pocket, which gains affinity for m^7GpppG cap analogues (or eventually for other ligands) only in the presence of the guide-strand RNA (and vice versa). It is thus tempting to speculate that the second sulphate ion observed in the context of the *Nc* QDE-2 MID domain could occupy such a second ligand-binding site.

Ligand binding by the two sites is expected to be cooperative. Accordingly, *Dm* AGO1b binding to miRNAs, m^7GpppG cap analogue and GW182 protein was abrogated by substituting the invariant tyrosine residue in the first binding site (*Dm* AGO1b residue Tyr 619, corresponding to *Nc* QDE-2 Tyr 595) or swapping the charge of an aspartic residue in the second ligand-binding site (*Dm* AGO1b residue Asp 627, corresponding to *Nc* QDE-2 Asp 603; Djuranovic *et al*, 2010). The inhibitory effect of the D627K substitution on miRNA binding is surprising because this residue is distant from the miRNA-binding site (Djuranovic *et al*, 2010). The equivalent residue in the *Nc* QDE-2 MID domain Asp 603 is located less than 7 Å away from the second sulphate (Fig 3A,B). Consequently, the Asp 627 side chain of *Dm* AGO1b could indeed participate in binding a ligand that superimposes with the second sulphate.

DISCUSSION

The structure of the *Nc* QDE-2 MID domain suggests that AGO proteins could be regulated by positive cooperativity of distinct ligands. Indeed, the structure reveals two precisely coordinated sulphate ions in adjacent binding pockets. Two highly conserved lysines (Lys 599 and Lys 638) coordinate both sulphates simultaneously and are thus shared by both binding pockets. Clearly, the first of the two sulphates on the *Nc* QDE-2 MID domain occupies the binding pocket for the 5′-terminal phosphate of the miRNA/short interfering RNA. The second sulphate is coordinated directly by three of the six most highly conserved side chains in AGO MID domains (that is, T(S)610, Lys 599 and Lys 638; Fig 1D). These residues are conserved not only in sequence but also in their structural orientation in bacterial and archaeal MID domains (supplementary Fig S1A,B online). Thus, the second sulphate might mark an ancient second (allosteric) binding site that has evolved different specificities in different AGO homologues (for example, for the cap structure, GW182 or other ligands).

We cannot predict what the ligand for the second binding site might be for different AGOs, because the properties and accessibility of this second binding site would depend on the orientation and the nature of the side chains provided by the PIWI domain. Indeed, in the current structures of the archaeal and eubacterial proteins, the PIWI domain restricts access to the second ligand-binding site. Nevertheless, the situation might be different for the eukaryotic AGOs, as Djuranovic *et al* (2010) observed an allosteric effect between the two binding sites using protein constructs that contained the PIWI domain. For AGO proteins involved in the miRNA pathway, the second ligand-binding site was suggested to bind to the m^7GpppN cap structure, or residues from the GW182 protein (Djuranovic *et al*, 2010). In agreement with this hypothesis, substituting residues in either of the two binding pockets simultaneously abrogated *Dm* AGO1 binding to miRNAs, GW182 and m^7GpppG cap analogue (Djuranovic *et al*, 2010). However, other AGO proteins, such as *Nc* QDE-2, might have evolved to be regulated by distinct ligands or might no longer be regulated by a second ligand. Nevertheless, for those AGO proteins that show ligand-dependent regulation, the coupled participation of the invariant lysines, Lys 599 and Lys 638, in both ligand-binding sites would be an attractive and mechanistically simple explanation for the ligand-dependent regulation of AGO protein function.

METHODS

Cloning, protein expression, purification and crystallization. The sequence encoding the MID domain of *Nc* QDE-2 was amplified by PCR from a pBluescript-SK^+ plasmid containing the genomic sequence of the *qde-2* gene (provided by G. Macino) and subsequently inserted into the *Nco*I and *Not*I sites of

the pETM60 vector (derived from pET24d; Novagen) using the primer pairs: 5′-CATGCCATGGCAGTCAAGGTCGCCAGA CCTT-3′ (forward) and 5′-ATAAGAATGCGGCCGCTTAGATATTG TGATTGGTGCCGC-3′ (reverse).

The resulting NusA–6 × His–MID protein fusion was expressed in the *Escherichia coli* BL21 Star (DE3) strain (Invitrogen) at 20 °C overnight. Protein expression was induced with 1 mM isopropyl β-D-1-thiogalactopyranoside at an A_{600} of 0.6. The fusion protein was purified first by a Ni^{2+}-affinity step. The removal of the NusA and polyhistidine tags was done by proteolytic digestion using tobacco etch virus protease at 4 °C overnight. The MID domain was purified further by heparin affinity chromatography and subsequent gel filtration (HiLoad 26/60 Superdex 75 pg; GE Healthcare). Finally, the QDE-2 MID domain was concentrated to 25 mg per ml in 10 mM HEPES (pH 7.2), 150 mM NaCl and 1 mM DTT.

Crystalline clusters of native and selenomethionine-substituted QDE-2 MID domain were grown by hanging drop vapour diffusion over a 500 µl reservoir at 18 °C. The protein solution described above was mixed 1:1 with various solutions containing 16–30% polyethylene glycol 4000 and 150–250 mM ammonium sulphate. Crystals were optimized by microseeding and flash-frozen in liquid nitrogen in the respective reservoir solution supplemented with 10 or 15% glycerol for cryoprotection.

Data collection, structure solution and refinement. Diffraction data were recorded on a PILATUS 6M detector (Broennimann *et al*, 2006) at the beamline PXII of the Swiss Light Source at a wavelength of 1.0 Å for the native data set and at the absorption peak of the selenomethionine derivative at 0.9792 Å. Diffraction images were processed with XDS (Kabsch, 1993). The crystals belong to space group $C222_1$, with one molecule per asymmetric unit and 52% solvent. The structure was solved from the selenomethionine data by single anomalous dispersion. AutoSHARP (Vonrhein *et al*, 2007) was used to search for five selenium sites per molecule. The assignment of the correct hand and solvent flattening was done automatically. ARP/wARP (Cohen *et al*, 2008) built a partial model comprising 99% of the backbone and 93% of the side chains docked correctly. The model was completed manually in COOT (Emsley & Cowtan, 2004) by using the native data. Refinement was done using Phenix (Adams *et al*, 2010) and COOT iteratively. The structure was validated with MOLPROBITY (Davis *et al*, 2007) and WHATCHECK (Hooft *et al*, 1996).

Accession codes. The coordinates of the MID domain have been deposited at the PDB with the accession code 2xdy.

Supplementary information is available at *EMBO reports* online (http://www.emboreports.org).

ACKNOWLEDGEMENTS

We thank G. Macino for providing *Neurospora crassa* QDE-2 plasmid DNA and R. Büttner for technical assistance. We thank the staff at the PX beamlines of the Swiss Light Source for assistance with data collection. This study was supported by the Max Planck Society, by grants from the Deutsche Forschungsgemeinschaft (FOR855 and the Gottfried Wilhelm Leibniz Program awarded to E.I.), and by the Sixth Framework Programme of the European Commission through the Silencing RNAs: Organisers and Coordinators of Complexity in Eukaryotic Organisms (SIROCCO) Integrated Project LSHG-CT-2006-037900.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

REFERENCES

Adams PD *et al* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66:** 213–221

Broennimann C *et al* (2006) The PILATUS 1M detector. *J Synchrotron Radiat* **13:** 120–130

Cohen SX *et al* (2008) ARP/wARP and molecular replacement: the next generation. *Acta Crystallogr D Biol Crystallogr* **64:** 49–60

Davis IW *et al* (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* **35:** W375–W383

Djuranovic S, Zinchenko MK, Hur JK, Nahvi A, Brunelle JL, Rogers EJ, Green R (2010) Allosteric regulation of Argonaute proteins by miRNAs. *Nat Struct Mol Biol* **17:** 144–150

Emsley P, Cowtan K (2004) COOT: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60:** 2126–2132

Eulalio A, Huntzinger E, Izaurralde E (2008a) Getting to the root of miRNA-mediated gene silencing. *Cell* **132:** 9–14

Eulalio A, Huntzinger E, Izaurralde E (2008b) GW182 interaction with Argonaute is essential for miRNA-mediated translational repression and mRNA decay. *Nature Struct Mol Biol* **15:** 346–353

Fulci V, Macino G (2007) Quelling: post-transcriptional gene silencing guided by small RNAs in *Neurospora crassa*. *Curr Opin Microbiol* **10:** 199–203

Hooft RW, Vriend G, Sander C, Abola EE (1996) Errors in protein structures. *Nature* **381:** 272

Jínek M, Doudna JA (2009) A three-dimensional view of the molecular machinery of RNA interference. *Nature* **457:** 405–412

Kabsch W (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J Appl Crystallogr* **26:** 795–800

Kinch LN, Grishin NV (2009) The human Ago2 MC region does not contain an eIF4E-like mRNA cap binding motif. *Biol Direct* **4:** 2

Kiriakidou M, Tan GS, Lamprinaki S, De Planell-Saguer M, Nelson PT, Mourelatos Z (2007) An mRNA m7G cap binding-like motif within human Ago2 represses translation. *Cell* **129:** 1141–1151

Lingel A, Simon B, Izaurralde E, Sattler M (2003) Structure and nucleic-acid binding of the *Drosophila* Argonaute 2 PAZ domain. *Nature* **426:** 465–469

Ma JB, Yuan YR, Meister G, Pei Y, Tuschl T, Patel DJ (2005) Structural basis for 5′-end-specific recognition of guide RNA by the *A. fulgidus* PIWI protein. *Nature* **434:** 666–670

Marcotrigiano J, Gingras AC, Sonenberg N, Burley SK (1997) Cocrystal structure of the messenger RNA 5′ cap-binding protein (eIF4E) bound to 7-methyl-GDP. *Cell* **89:** 951–961

Parker JS, Roe SM, Barford D (2004) Crystal structure of a PIWI protein suggests mechanisms for siRNA recognition and slicer activity. *EMBO J* **23:** 4727–4737

Parker JS, Roe SM, Barford D (2005) Structural insights into mRNA recognition from a PIWI domain–siRNA guide complex. *Nature* **434:** 663–666

Parker JS, Parizotto EA, Wang M, Roe SM, Barford D (2009) Enhancement of the seed-target recognition step in RNA silencing by a PIWI/MID domain protein. *Mol Cell* **33:** 204–214

Song JJ, Liu J, Tolia NH, Schneiderman J, Smith SK, Martienssen RA, Hannon GJ, Joshua-Tor L (2003) The crystal structure of the Argonaute2 PAZ domain reveals an RNA binding motif in RNAi effector complexes. *Nat Struct Mol Biol* **10:** 1026–1032

Song JJ, Smith SK, Hannon GJ, Joshua-Tor L (2004) Crystal structure of Argonaute and its implications for RISC slicer activity. *Science* **305:** 1434–1437

Tolia NH, Joshua-Tor L (2007) Slicer and the Argonautes. *Nat Chem Biol* **3:** 36–43

Vonrhein C, Blanc E, Roversi P, Bricogne G (2007) Automated structure solution with autoSHARP. *Methods Mol Biol* **364:** 215–230

Wang Y, Sheng G, Juranek S, Tuschl T, Patel DJ (2008a) Structure of the guide-strand-containing Argonaute silencing complex. *Nature* **456:** 209–213

Wang Y, Juranek S, Li H, Sheng G, Tuschl T, Patel DJ (2008b) Structure of an Argonaute silencing complex with a seed-containing guide DNA and target RNA duplex. *Nature* **456:** 921–926

Wang Y, Juranek S, Li H, Sheng G, Wardle GS, Tuschl T, Patel DJ (2009) Nucleation, propagation and cleavage of target RNAs in Ago silencing complexes. *Nature* **461:** 754–761

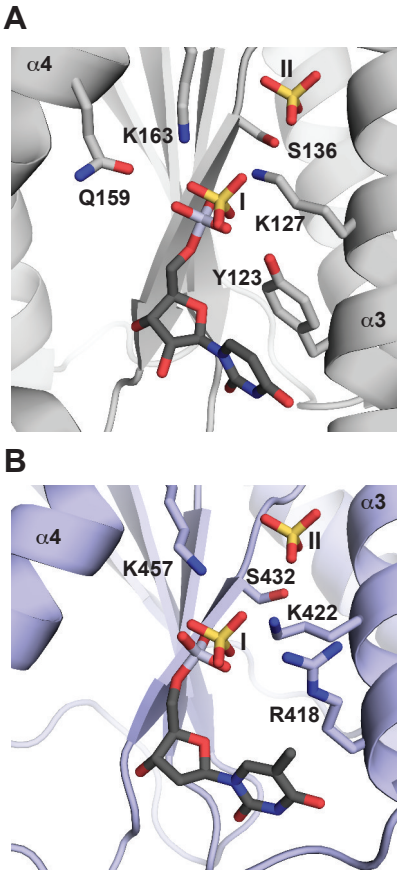
Yan KS, Yan S, Farooq A, Han A, Zeng L, Zhou MM (2003) Structure and conserved RNA binding of the PAZ domain. *Nature* **426:** 468–474

Yuan YR, Pei Y, Ma JB, Kuryavyi V, Zhadina M, Meister G, Chen HY, Dauter Z, Tuschl T, Patel DJ (2005) Crystal structure of *A. aeolicus* Argonaute, a site-specific DNA-guided endoribonuclease, provides insights into RISC-mediated mRNA cleavage. *Mol Cell* **19:** 405–419

Supplementary information

Crystal structure and ligand binding of the MID domain of a eukaryotic Argonaute protein

Andreas Boland, Felix Tritschler, Susanne Heimstädt, Elisa Izaurralde and Oliver Weichenrieder



Supplementary Fig 1 Structural conservation of ligand-binding residues. The two sulfate ions (I and II) from the *Nc* QDE-2 MID domain (Fig 2B, same view) superimposed onto the guide RNA 5' nucleotide-binding pocket of AGO proteins from *Archaeoglobus fulgidus* (A) and *Thermus thermophilus* (B). Only the MID domains and the 5' nucleotide of the co-crystallized nucleic acid substrates are shown. (Red: oxygen; blue: nitrogen; yellow: sulfur; violet: phosphorus). (**A**) *Archaeoglobus fulgidus* AGO MID domain (PDB-ID: 2BGG, Parker *et al*, 2005) together with the 5' nucleotide of the guide RNA. (**B**) *Thermus thermophilus* AGO MID domain (PDB-ID: 3HVR, Wang *et al*, 2009) together with the 5' nucleotide of the guide DNA.

Table S1. Data collection, phasing and refinement statistics.

	Native	**Selenomethionine substituted**
Data collection		
Space group	$C222_1$	$C222_1$
Unit cell (Å, °)	80.1, 91.39, 42.93 90, 90, 90	80.17, 91.52, 43.16 90, 90, 90
Molecules per asymmetric unit	1	1
Resolution range (Å) [a]	60.238-2.2 (2.26-2.2)	60.302-2.2 (2.26-2.2)
Unique reflections	8285	8259
Completeness, (%) [a]	99.4 (100)	
I/σ(I) [a]	18.36 (2.94)	8.44 (1.90)
R_{merge} (%) [a]	4.7 (49.6)	7.9 (55.1)
Average redundancy [a]	3.6 (3.6)	
Phasing Statistics		
Phasing power (anom.)		1.078
R_{cullis} (anom.)		0.809
Mean FOM		0.29525
Anomoulous Completeness (%) [a]		98.5 (96.1)
Anomoulous average redundancy [a]		1.8 (1.7)
Structure refinement		
Resolution range (Å) [a]	60.238-2.2 (2.26-2.2)	
R-factor / R-free (%)	19.29 / 22.29	
Number of		
protein atoms	989	
sulfate ions	2	
glycerol molecules	1	
water molecules	52	
Ramachandran plot [b]		
most favoured (%)	98.4	
additional allowed (%)	1.6	
outliers (%)	0	
RMSD bond lengths, (Å)	0.004	
RMSD bond angles, (°)	0.776	

[a] Numbers in parentheses are for the highest-resolution shell

[b] Calculated using Molprobity

Crystal structure of the MID-PIWI lobe of a eukaryotic Argonaute protein

Andreas Boland, Eric Huntzinger, Steffen Schmidt, Elisa Izaurralde[1], and Oliver Weichenrieder[1]

Max Planck Institute for Developmental Biology, Spemannstrasse 35, D-72076 Tübingen, Germany

Edited by Jennifer A. Doudna, University of California, Berkeley, CA, and approved May 6, 2011 (received for review March 10, 2011)

Argonaute proteins (AGOs) are essential effectors in RNA-mediated gene silencing pathways. They are characterized by a bilobal architecture, in which one lobe contains the N-terminal and PAZ domains and the other contains the MID and PIWI domains. Here, we present the first crystal structure of the MID-PIWI lobe from a eukaryotic AGO, the *Neurospora crassa* QDE-2 protein. Compared to prokaryotic AGOs, the domain orientation is conserved, indicating a conserved mode of nucleic acid binding. The PIWI domain shows an adaptable surface loop next to a eukaryote-specific α-helical insertion, which are both likely to contact the PAZ domain in a conformation-dependent manner to sense the functional state of the protein. The MID-PIWI interface is hydrophilic and buries residues that were previously thought to participate directly in the allosteric regulation of guide RNA binding. The interface includes the binding pocket for the guide RNA 5′ end, and residues from both domains contribute to binding. Accordingly, micro-RNA (miRNA) binding is particularly sensitive to alteration in the MID-PIWI interface in *Drosophila melanogaster* AGO1 in vivo. The structure of the QDE-2 MID-PIWI lobe provides molecular and mechanistic insight into eukaryotic AGOs and has significant implications for understanding the role of these proteins in silencing.

Proteins of the Argonaute (AGO) family play essential roles in RNA-mediated gene silencing mechanisms in eukaryotes (1, 2). They are loaded with small noncoding RNAs to form the core of RNA-induced silencing complexes, which repress the expression of target genes at the transcriptional or posttranscriptional level (1, 2). The targets to be silenced are selected through base-pairing interactions between the loaded small RNA (also known as the guide RNA) and an mRNA target containing partially or fully complementary sequences (1–3).

Thus far, structural information on full-length AGOs has been available only for the homologous proteins from *Archaea* and *Eubacteria*, which preferentially use DNA as a guide (4–10). These studies revealed that AGOs consist of four domains: the N-terminal domain; the PAZ domain, which binds the 3′ end of guide RNAs/DNAs; the MID domain, which provides a binding pocket for the 5′ phosphate of guide RNAs/DNAs; and the PIWI domain, which adopts an RNase H fold and has endonucleolytic activity in some, but not all, AGOs (4–11).

For the eukaryotic AGO clade of Argonaute proteins, structural information is available only for the isolated PAZ domains of *Drosophila melanogaster* (*Dm*) AGO1 and AGO2, human AGO1 (12–16) and the MID domains of human AGO2 and *Neurospora crassa* (*Nc*) QDE-2 (17, 18). Structural information is also available for PAZ domains of the PIWI clade of AGOs (19, 20). These studies showed that the PAZ and MID domains of eukaryotic AGOs adopt folds similar to the prokaryotic homologs and recognize the 3′- and 5′-terminal nucleotides of the guide strand, respectively, in a similar manner to their prokaryotic counterparts (12–18).

Our previous structure of the isolated *Nc* QDE-2 MID domain revealed that the 5′-nucleotide binding site shares residues with a second, adjacent sulfate ion-binding site, suggesting that the 5′-terminal nucleotide of the guide RNA and a second ligand may bind cooperatively to the MID domain of eukaryotic AGOs (17). These findings supported the observation of Djuranovic et al. (21) that the isolated MID domains of certain eukaryotic AGOs [i.e., those involved in the micro-RNA (miRNA) pathway] contain a second, allosteric nucleotide binding site with an affinity for m^7GpppG cap analogs. However, considering the structures of the prokaryotic AGOs, it was unclear whether the putative second ligand-binding site would be accessible in the presence of the PIWI domain.

To address this question and gain further molecular insight into eukaryotic AGOs, we determined the crystal structure of the entire MID-PIWI lobe of the *Nc* QDE-2 protein, tested its RNA-binding properties in vitro, and analyzed its implications in vivo in the context of the *Dm* AGO1 protein. The structure provides a detailed, high-resolution view of the MID-PIWI interface in a eukaryotic AGO protein and shows that the two domains are oriented very similarly to their prokaryotic counterparts, indicating a conserved mode of guide RNA/DNA strand recognition. However, despite these similarities, the PIWI-domain exhibits eukaryote-specific structural features that might act as sensors for the functional state of the protein. Finally, we show that residues that have been implicated in allosteric regulation in previous studies (21) are in fact involved in MID-PIWI interdomain interactions and contribute to guide RNA binding by stabilizing the MID-PIWI interface.

Results

The Eukaryotic AGO MID-PIWI Lobe Adopts a Fold Highly Similar to the Prokaryotic Homologs. The *Neurospora crassa* Argonaute protein QDE-2 (*Nc* QDE-2) is a close sequence homolog of eukaryotic AGOs that act in the small interfering RNA (siRNA) and miRNA pathways [e.g., sequence identities with Hs AGO2 and Dm AGO1 are 30% and 29.7%, respectively (Fig. S1) (22, 23)]. We obtained diffracting crystals of a QDE-2 fragment containing the MID and PIWI domains (amino acids 506–938) and determined the structure at 3.65 Å resolution (crystal form I; Table S1). The structure revealed a disordered loop (loop L3, Fig. 1*A*, and Fig. S1) including amino acids K786–A840 of the PIWI domain, which we replaced with a Gly-Ser (GSG) linker to generate crystals of a MID-PIWI ΔL3 protein that diffracted to 1.85 Å resolution. This structure (crystal form II) was refined to an R_{work} of 19.6% (R_{free} = 23.6%), whereas crystal form I yielded an R_{work} of 23.3% (R_{free} = 25.2%) (Table S1).

This structure reveals the details of a eukaryotic AGO MID-PIWI lobe. The individual MID and PIWI domains superimpose well with the structures of previously determined archaeal and eubacterial AGO MID and PIWI domains (Fig. 1*A* vs. 1*B* and

Author contributions: A.B., E.I., and O.W. designed research; A.B., E.H., and O.W. performed research; A.B., E.H., S.S., E.I., and O.W. analyzed data; and A.B., E.I., and O.W. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: The accession codes and coordinates of the QDE-2 MID-PIWI lobe and MID-PIWI ΔL3 have been deposited in the Protein Data Bank, www.pdb.org (PDB ID codes 2yhb and 2yha).

[1]To whom correspondence may be addressed. E-mail: elisa.izaurralde@tuebingen.mpg.de or oliver.weichenrieder@tuebingen.mpg.de.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1103946108/-/DCSupplemental.

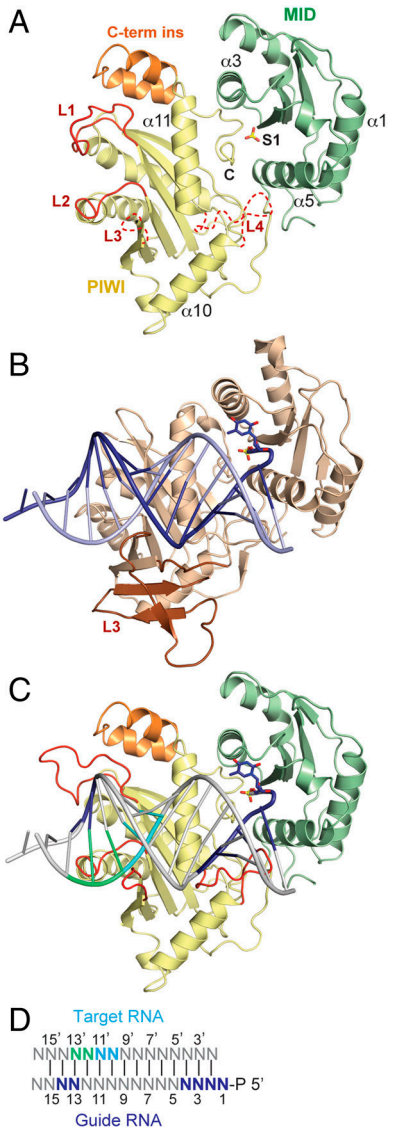


Fig. 1. Structure of the *Neurospora crassa* QDE-2 MID-PIWI lobe. (*A*) Ribbon representation of the QDE-2 MID-PIWI ΔL3 structure showing the position of a bound sulfate ion (S1) as sticks (red, oxygen; yellow, sulfur). Disordered portions of the polypeptide chains are indicated with dashed lines. Selected loops and secondary structure elements are labeled. Loop L1 is shown in conformation II. Loop L3 was deleted and replaced by a GSG linker. Loop L4 is modeled based on prokaryotic structures [PDB ID code 1W9H (11)]. The eukaryotic-specific insertion is colored orange. (*B*) Ribbon representation of the MID-PIWI lobe of *Thermus thermophilus (Tt)* AGO in complex with a guide DNA-target RNA duplex, generated from PDB ID code 3HJF (10). (*C*, *D*) Model for a nucleic acid bound to the QDE-2 MID-PIWI lobe, based on the superposition with the *Tt* PIWI domain (PDB ID code 3HJF, as shown in panel *B*). Specific bases of the guide DNA and target RNA strands are colored as indicated in panel *D*. Loop L1 is shown in conformation I.

Fig. S2 *A* and *B*). Importantly, the relative orientation of the MID and PIWI domains is similar to the prokaryotic proteins, with the C-terminal residues (labeled "C" in Fig. 1*A*) of the protein deeply inserted into the MID-PIWI interface. Therefore, this structure can be superimposed over bacterial AGO-nucleic acid complexes to identify structural features that are specific to eukaryotes (Fig. 1 *A*–*D*).

The structure of the *Nc* QDE-2 MID-PIWI lobe serves as a prototype for MID-PIWI domains from all the eukaryotic clades of the Argonaute protein family, including the AGO, PIWI, and WAGO clades (Fig. S1 and ref. 1). However, we limit our analysis here to the AGO clade, which contains *Nc* QDE-2 and eukaryotic AGOs involved in the siRNA and miRNA pathways.

The PIWI Domain and the Interaction with Target mRNA. The PIWI domain of *Nc* QDE-2 (residues H644–I938) adopts an RNase H fold with a catalytically active DDD motif, which is less common than the DDH motif present in most eukaryotic AGOs (Fig. S1). A structure-based alignment (Fig. S1) identified a series of loops on the putative nucleic acid-binding surface of the PIWI domain [Fig. 1*A*, loops L1 (H667–P681), L2 (G746–Q751), L3 (K788–A840), and L4 (F873–I882)]. Their significance is best understood in the context of a double stranded nucleic acid substrate that can be placed on the QDE-2 MID-PIWI lobe by the structural superposition with *Thermus thermophilus* AGO-nucleic acid complex [Fig. 1 *B* and *C*; Protein Data Bank (PDB) ID code 3HJF; ref. 10]. In this particular model substrate, the 3′ end of the guide DNA strand is not anchored in the PAZ domain. Instead, nucleotides 2 to 15 of the guide DNA are base-paired to an RNA target strand, forming a duplex extending beyond the seed sequence (see Fig. 1*D* for a numbering of the respective nucleotides), which is typical for siRNA targets but rather rare for animal miRNA targets (3).

The scissile bond of the model target strand (between nucleotides 10′ and 11′) fits nicely in the catalytic site (Fig. 2 *A* and *B*), and loop L2 is perfectly positioned to probe the minor groove of the duplex (base pair 13) with S748 (Fig. 2 *C* and *D*). Loop L2 is highly conserved in all eukaryotic AGOs and likely fixes the phosphodiester backbone of the target strand (nucleotides 12′ and 13′) via E749 and nucleotide 14 from the guide strand via Q751 (Fig. 2 *C* and *D*) in cases where the downstream duplex is formed (e.g., in the case of fully complementary targets).

Loop L3 is disordered in crystal form I and was deleted in crystal form II. Judging from the prokaryotic AGO structures, the central parts of this loop likely organize the missing lobe (containing the N-term and PAZ domains) of the AGO protein (Fig. S2*B*), whereas the N-terminal residues of loop L3 [including the conserved K786 (Fig. S1)] might assist loop L2 in fixing the backbone of the target strand (nucleotides 10′ and 11′). Therefore, regardless of whether the downstream duplex is formed, all eukaryotic AGO proteins are likely to bind the target strand between loops L2 and L3 in a conserved way.

The most interesting feature in this context is loop L1, which changes conformation between crystal form I (conformation I, Fig. 1*C*) and crystal form II (conformation II, Fig. 1*A* and Fig. S1) and is more variable in sequence and length than loop L2. In conformation II, loop L1 clashes with the base-paired guide strand at nucleotides 12–14 (Fig. 2*D*), whereas conformation I may recognize and stabilize nucleotides 13 and 14 in the duplex via H667 (Fig. 2*C*). Consequently, this "switch loop" could sense or control the formation of a downstream duplex, with important consequences for subsequent steps in the siRNA and miRNA pathways (e.g., recruitment of GW182 proteins). A switch in loop L1 upon duplex formation has previously been described for *Tt* AGO (10), indicating that this may indeed be an important and conserved function in AGO proteins. In contrast to the *Tt* AGO structure, however, we do not observe a correlated flip of the beta strand β6 (10).

Finally, loop L4 is disordered in this structure (note that in Figs. 1 and 2 this loop is modeled as in prokaryotic structures). Based on prokaryotic structures, loop L4 likely helps to fix nucleotides 2 to 5 from the guide RNA seed region, even in the absence of a target strand, ensuring it is held in a hybridization-competent state for target seed recognition, as proposed for the prokaryotic homologs (6, 24) and *Caenorhabditis elegans* ALG1 (25).

The Eukaryotic PIWI Domains Contain a C-Terminal Insertion. A peculiar feature of the eukaryotic AGO clade is a C-terminal insertion

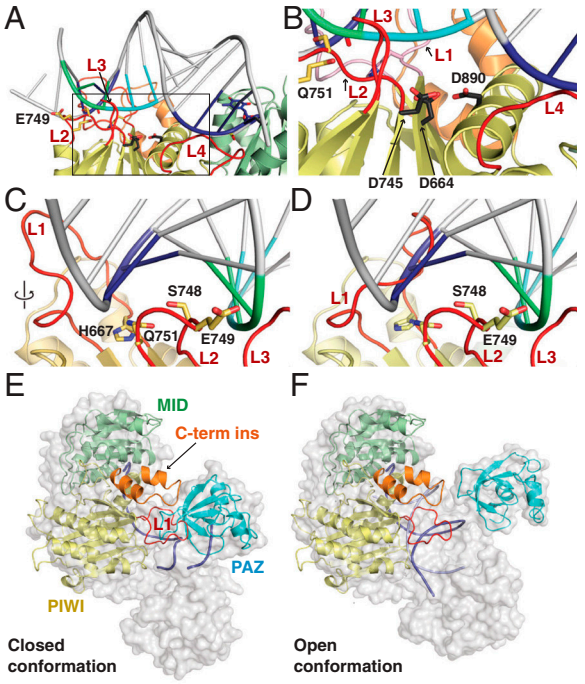


Fig. 2. Details of the nucleic acid binding surface. (*A*) View of the QDE-2 PIWI domain bound to a model guide DNA-target RNA duplex color-coded as in Fig. 1*D*. (*B*) Expanded view of the framed region in panel *A* showing the catalytic site of *Nc* QDE-2 PIWI domain and the position of the backbone phosphate linking the 10′-11′ bases of the target RNA relative to the catalytic residues (D664, D745, and D890). For clarity, loop L1 is colored pink. (*C*, *D*) Extended view of loops L1 and L2 in contact with the guide DNA-target RNA duplex. Panel (*C*) shows loop L1 in conformation I whereas panel *D* shows conformation II, in which loop L1 clashes with the duplex. (*E*, *F*) Relative orientation of eukaryotic AGO domains (ribbons), using bacterial AGO structures as template (transparent surfaces). (*E*) Closed conformation in the absence of a target strand, where the 3′ end of the guide strand is anchored in the PAZ domain [based on PDB ID code 3DLH (8)]. (*F*) Open conformation in the presence of a target strand, where the 3′ end of the guide strand is released [based on PDB ID code 3HJF (10)]. Note that in the closed conformation (*E*), the C-terminal insertion (C-term ins) and loop L1 of the QDE-2 MID-PIWI lobe can contact the PAZ domain. In both panels, loop L1 is shown in the conformation observed in crystal form I. However, in the closed conformation (*E*) loop L1 could adopt conformation II or any other similar conformation and still contact the PAZ domain.

(K901–G925) that, in QDE-2, folds into two closely packed helices (Fig. 1 *A* and *C*; in orange). This eukaryote-specific α-helical insertion is located right next to the switch loop L1, and also shows a minor conformational difference between the two crystal forms in the short turn connecting the helices. In contrast to loop L1, however, the α-helical insertion does not change position between the two crystal forms, and there are no specific contacts to loop L1 (Fig. 1 *A* vs. *C*). Judging from the superposition with a closed conformation of *Tt* AGO (PDB ID code 3DLH; ref. 8), the α-helical insertion together with loop L1 is able to contact the PAZ domain (Fig. 2*E*). The respective *Tt* AGO structure contains a guide strand with the 5′ and 3′-terminal nucleotides anchored in their specific binding pockets in the MID and PAZ domains, respectively. In contrast, in the superposition with an open conformation of *Tt* AGO bound to a guide-target duplex, in which the 3′-terminal nucleotide of the guide RNA is released from the PAZ domain (PDB ID code 3HJF; ref. 10), the PAZ domain moves away from loop L1 and the α-helical insertion (Fig. 2*F*). These observations suggest that loop L1 and the α-helical insertion establish conformation-dependent interactions with the PAZ domain in eukaryotic AGOs. These structural elements might, therefore, act as a sensor for the functional state of the protein and play a regulatory role.

The Binding Pocket for the 5′ End of the Guide Strand Forms at the MID-PIWI Interface. The structure of the MID domain (V506–N643), with its Rossmann-like fold, remains highly similar to the previously determined structure in isolation (17, 18). Accordingly, we also find a sulfate ion (S1) in the binding pocket of the guide RNA 5′ end, coordinated by the highly conserved residues Y595, K599, and K638 (Figs. 1*A* and 3*A*). This sulfate ion mimics and occupies the position of the 5′ phosphate that is characteristic of siRNAs and miRNAs. Importantly, when we superimpose the prokaryotic nucleic acid substrates, using the PIWI domain as a reference, the 5′ phosphate of the guide strand is at a distance of less than 2.5 Å from the sulfate, and the first nucleotide stacks on Y595, as previously suggested (Fig. 3*B* and refs. 5, 17, and 18). This confirms that the relative orientation of the MID and PIWI domains and the geometry of substrate binding are very similar between prokaryotes and eukaryotes, validating the interpretations of the prokaryotic complexes in this respect.

Indeed, as in prokaryotes, the binding pocket for the 5′ phosphate of the guide strand lies at the MID-PIWI interface and is completed by the C-terminal carboxyl group of the protein (labeled "C" in Figs. 1*A* and 3 *A* and *B*). Furthermore, R895 (PIWI), which is present in most eukaryotic proteins from the AGO and WAGO clades, also participates in the 5′ binding pocket (Fig. 3 *A* and *B*). It forms hydrogen bonds to Y595 (MID) and

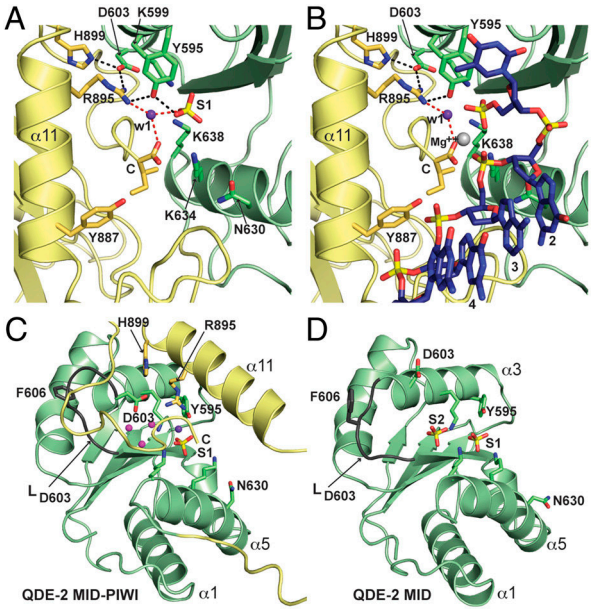


Fig. 3. Details of the MID-PIWI interface. (*A*, *B*) Extended view of the guide strand 5′-terminal nucleotide-binding pocket, without (*A*) and with (*B*) bound guide strand (DNA) model from the superposition of the *Tt* MID-PIWI domain in complex with a guide-target duplex (Fig. 1*C*). The guide strand is shown as sticks colored according to the scheme in Fig. 1*D*. The 5′ base of the guide strand stacks onto Y595, whereas the 5′ phosphate superimposes with sulfate S1. Relevant side chains are shown as sticks. Hydrogen bonds are shown as dotted lines (red for water w1). The water molecule (w1) is shown in purple, and a magnesium ion from the *Tt* AGO structure (PDB ID codes 3DLH and 3HJF) is shown in gray (red, oxygen; blue, nitrogen; yellow, sulfur and phosphorus). (*C*, *D*) Comparison of QDE-2 MID domain structures in the context of the MID-PIWI lobe (*C*) or in isolation (*D*). Loop L_{D603} (dark gray) changes conformation between the two structures. The structure of the isolated QDE-2 MID domain contains two bound sulfate ions (S1 and S2). S1 marks the position of the 5′ phosphate of the guide strand, whereas S2 marks the position of a putative second ligand-binding site [PDB ID code 2xdy (17)]. In the structure of the QDE-2 MID-PIWI lobe, the S2 binding site is occluded by the C-terminal tail of the protein, and S2 is replaced by three water molecules (magenta). Relevant side chains and secondary structure elements are labeled. D603 corresponds to the aspartate proposed by Djuranovic et al. (21) to mediate allosteric regulation of miRNA binding. MID, pale green; PIWI, yellow.

to an important water molecule (w1, purple sphere in Fig. 3 *A* and *B*) that contacts both the C-terminal carboxyl group and the bound sulfate. In the presence of the guide strand, the carboxyl group is also expected to coordinate a magnesium ion (Fig. 3*B*, gray sphere, 3DLH and 3HJF, and refs. 8 and 10) and to bend the guide RNA backbone between nucleotides 1 and 2 to a degree similar to that observed in the prokaryotic complexes (5, 6, 8, 10). Consequently, whereas the phosphates of nucleotide 1 and 3 would be bridged by the magnesium ion, the ribose of nucleotide 2 would be contacted by the conserved N630 (MID), the phosphate of nucleotide 3 by K634 (MID), and the phosphate of nucleotide 4 by Y887 (PIWI) (Fig. 3 *A* and *B*). This places nucleotides 2–4 from the guide RNA seed region in a hybridization-competent state (Fig. 3*B*), that is likely to be stabilized further by presently disordered residues from loop L4 (see above).

Interestingly, in the PIWI clade of the AGO protein family, the highly conserved residues N630 and K634 are replaced by lysine and glutamine, respectively, and in the WAGO clade, the otherwise invariant Y595 is a histidine. These clade-specific diagnostic differences in the 5′ binding pocket are worth pointing out here and deserve further structural investigation.

The Hydrophilic MID-PIWI Interface. A comparison of the QDE-2 MID domain in the presence (Fig. 3*C*) or absence (Fig. 3*D* and ref. 16) of the PIWI domain revealed an interesting conformational difference in a loop between residues D603 and V608 (termed loop L_{D603}). This loop had previously been implicated in the allosteric regulation of guide RNA binding, and mutational analysis has suggested that an aspartate (D603 in *Nc* QDE-2) could mediate this effect (21); this residue is conserved in a subset of eukaryotic AGOs (Fig. S1). Furthermore, the isolated QDE-2 MID domain bound a second sulfate ion (S2) in a position that would be ideal for an allosteric ligand (Fig. 3*D*) and could be contacted by D603 if loop L_{D603} adopted the conformation observed in the human AGO2 MID domain (18).

The structure of the *Nc* QDE-2 MID-PIWI lobe now reveals that loop L_{D603} can indeed change into the conformation observed for the human MID domain (Fig. 3*C* and Fig. S3*A*), which creates space for the insertion of the C-terminal tail of the protein into the MID-PIWI interface (Fig. 3*C* and Fig. S3*A*). As a result, the tail displaces the sulfate ion (S2) and occludes the second sulfate ion-binding site in the present structure, deeply burying three water molecules (magenta spheres). These are coordinated by residues K599, H609, V608 (main chain carbonyl), T610, K638 and D603 from the MID domain, and R895 and Y936 (main chain carbonyl) from the PIWI domain, most of which are highly conserved (Fig. 3*C* and Figs. S1 and S3*B*). Importantly, D603 is now buried in the interface as well and directly contacts R895 from the PIWI domain, which is part of the 5′ guide RNA binding pocket (Fig. 3 *A* and *B* and Fig. S3*B*). Furthermore, D603 also contacts H899 from the PIWI domain, a residue that is conserved only among animal AGOs implicated in the miRNA pathway (Fig. 3 *A* and *B* and Fig. S3*B*).

The polar and hydrophilic character of the interface is conserved within the eukaryotic AGO and WAGO clades and explains why a mutation in D603 has long-range effects on guide RNA binding (21) without having to invoke allosteric control.

Guide RNA Binding by QDE-2 Requires the MID-PIWI Interaction. To investigate the contribution of the PIWI domain to 5′-nucleotide binding in vitro, we used size exclusion chromatography. As a substrate, we chose a 10-mer guide RNA mimic that started with a 5′-phosphorylated uridine (23). We first showed that the MID domain alone could not bind the substrate (Fig. 4*A*), whereas a protein fragment containing the MID and PIWI domains readily bound RNA, forming a 1:1 protein–RNA complex (Fig. 4*B*). This shows that the PIWI domain is indeed essential for guide RNA binding. Remarkably, the PIWI domain does not contribute to guide RNA binding only through residues that are part of the 5′-nucleotide binding pocket because the MID-PIWI ΔL3 protein from the high-resolution structure (crystal form II) exhibited reduced RNA-binding affinity (Fig. 4*C*). Therefore, we used the full-length MID-PIWI lobe for additional RNA-binding assays.

First, we showed that, in vitro, the guide RNA mimic binds specifically to the 5′-terminal nucleotide binding pocket because the substitution of the Y595 stacking platform with a leucine abolished binding, as expected (Fig. 4*D* and ref. 5). We also found guide RNA binding to be abolished by R895A (Fig. 4*E*), although R895 contacts the 5′ phosphate only indirectly via water w1 (Fig. 3*A*). Hence, guide RNA binding seems very sensitive to subtle alterations in the 5′ binding pocket. Additionally, the R895A substitution may also affect guide RNA binding via a destabilization of the MID-PIWI interface (Fig. 3*A* and Fig. S3*B*). This hypothesis is supported by an even more distant H899A mutation that also reduces RNA binding (Fig. 4*F*). Both R895 and H899 contact D603, located at the center of the interface (Fig. 3*A*). However, an additional D603K mutant could not be obtained in soluble form, indicating that the complete disruption of the interface strongly destabilizes the protein. Finally, we prepared a deletion mutant lacking the eukaryote-specific C-terminal

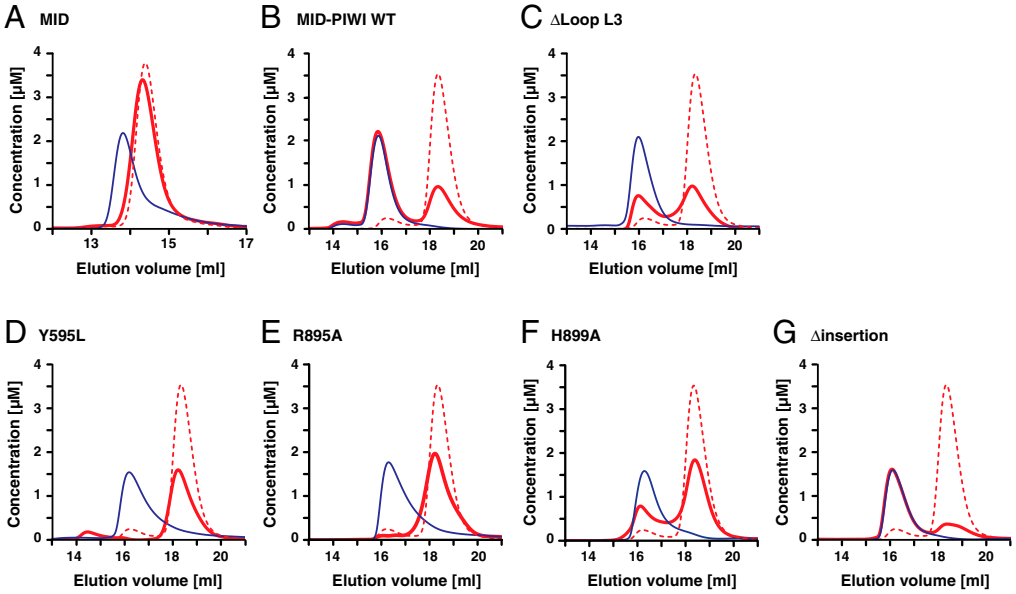


Fig. 4. Mutational analysis of the *Nc* QDE-2 MID-PIWI lobe. (*A*–*G*) A 5′-phosphorylated 10-nucleotide long RNA oligo with a 5′-terminal uridine was analyzed by size exclusion chromatography, either in the absence (dashed red lines) or presence (solid red lines) of the indicated proteins. Proteins analyzed in the presence of the RNA substrate are shown as solid blue lines. Note that the elution volume of the proteins does not change in the presence of nucleic acids. Concentrations were calculated from the relative absorption properties of the components.

insertion. This protein was soluble and bound the 10-mer 5′ guide RNA mimic (Fig. 4*G*). This indicates that the interface, including the very C-terminal end of the protein remains intact in this mutant and that the C-terminal insertion is, therefore, not required to bind to this particular substrate. This result is not surprising because prokaryotic AGOs bind guide nucleic acids even though they lack the C-terminal insertion (Fig. S1).

The MID-PIWI Domain Interface Is Required for *Dm* AGO1 to Bind miRNAs and GW182. To exploit the crystal structure in a physiological context and to gain additional functional insight into the miRNA pathway, we generated analogous mutations in the context of full-length, HA-tagged *Dm* AGO1 protein and tested them for interactions with an endogenous miRNA (Bantam) and the GW182 protein in *D. melanogaster* S2 cells. Previous studies have shown that GW182 proteins interact directly with AGOs and are essential for miRNA-mediated gene silencing in animal cells (26). Using a complementation assay (27), we also tested *Dm* AGO1 mutants for their ability to restore silencing in cells depleted of endogenous AGO1 (Fig. S4*A*).

In contrast to previous mutational analyses of human AGO2 and *Dm* AGO1 (27, 28), we could design and interpret our experiments in the context of the QDE-2 structure, which is 29.7% identical to *Dm* AGO1. Moreover, the hydrophilic MID-PIWI interface (including residues R895, D603, H899) is particularly well conserved in *Dm* AGO1 (Fig. S1). We generated three classes of mutations: (*i*) mutations on the putative RNA-binding interface, (*ii*) mutations in the MID-PIWI interface, and (*iii*) deletions of either the eukaryote-specific C-terminal insertion or of loop L1 (Table S2). A *Dm* AGO1 F2V2 mutant that is unable to bind miRNAs and GW182 (27, 29) served as a negative control.

Dm AGO1 mutants targeting the 5′-phosphate binding pocket were strongly impaired in miRNA binding. These include substitutions of the following residues (corresponding QDE-2 residues in brackets, Table S2): Y653L (Y595), K657E (K599), K690A (K634), and K694A (K638; Fig. 5 *A* and *C*). These mutants were also impaired in complementation assays, as expected (Fig. S4 *A*–*C*). Additionally, the replacement of loop L3 (*Dm* AGO1) with a glycine-serine linker strongly impaired miRNA binding (Fig. 5*B*; ΔLoop L3). This can be explained by a direct binding defect, as shown for QDE-2 (Fig. 4*C*), although it is possible that this deletion affects the function of the protein in a more severe manner, given that L3 also contacts the N-term-PAZ lobe (Fig. S2*B*). Accordingly, deleting loop L3 completely abolished silencing activity in complementation assays (Fig. S4 *A*–*C*). In contrast, mutations of residues in loop L2 that are predicted to contact the RNA in a target-bound state [E798 (E749) and Q800 (Q751), Fig. 2*C*] had no effect on miRNA binding (Fig. 5*C*, lanes 15 and 16). These mutants were also active in complementation assays (Fig. S4 *A*–*C*), indicating that the interaction of loop L2 with the target mRNA is dispensable, at least for partially complementary targets that are not sliced.

Importantly, mutations in the MID-PIWI interface impaired miRNA binding and silencing activity to different extents. These include D661K (D603), R937A (R895) and H941A (H899; Fig. 5 *A* and *B*; Fig. S4 *A*–*C*). These results show that the interactions observed in the QDE-2 structure are indeed conserved in *Dm* AGO1 and are crucial for miRNA binding. They also uncover a particular sensitivity of miRNA binding to the stability of the MID-PIWI interface and to the relative orientation of the two domains.

Unexpectedly, replacing the eukaryote-specific C-terminal insertion with a glycine-serine linker (Fig. 5*B*, lane 16; Table S2) also impaired miRNA binding and abolished silencing activity (Fig. S4 *A*–*C*). This is intriguing because it is unlikely that this insertion affects guide strand RNA binding directly and because the corresponding deletion in the context of QDE-2 did not affect the binding to a 10-mer guide RNA mimic in vitro (Fig. 4*G*). Hence, these observations may reflect interesting functional differences between AGOs that act in distinct pathways. Alternatively, the insertion may be important in an earlier step, such as during structural arrangements accompanying the loading of a full-length 21-mer miRNA with its 3′ end anchored in the PAZ domain. Indeed, in the context of a full-length *Dm* AGO1 protein the C-terminal insertion would be in a position to contact the PAZ domain (Fig. 2*E*).

Finally, we also tested the effect of deleting loop L1, which together with the C-terminal insertion contacts the PAZ domain. We generated a deletion mutant that completely lacks loop L1 (ΔL1+His) and a mutant lacking only the most distal tip of this loop (ΔL1) conserving residue H724 (H667). Histidine 667 may contribute to miRNA binding by contacting the guide strand downstream of the seed sequence (Fig. 2*C*). These mutants bound miRNA and GW182 and were fully active in complementation assays (Fig. S4 *A*–*D*). These results indicate that loop L1 is dispensable at least for miRNA-mediated silencing and further suggest that loop L1 and the C-terminal insertion play distinct functional roles.

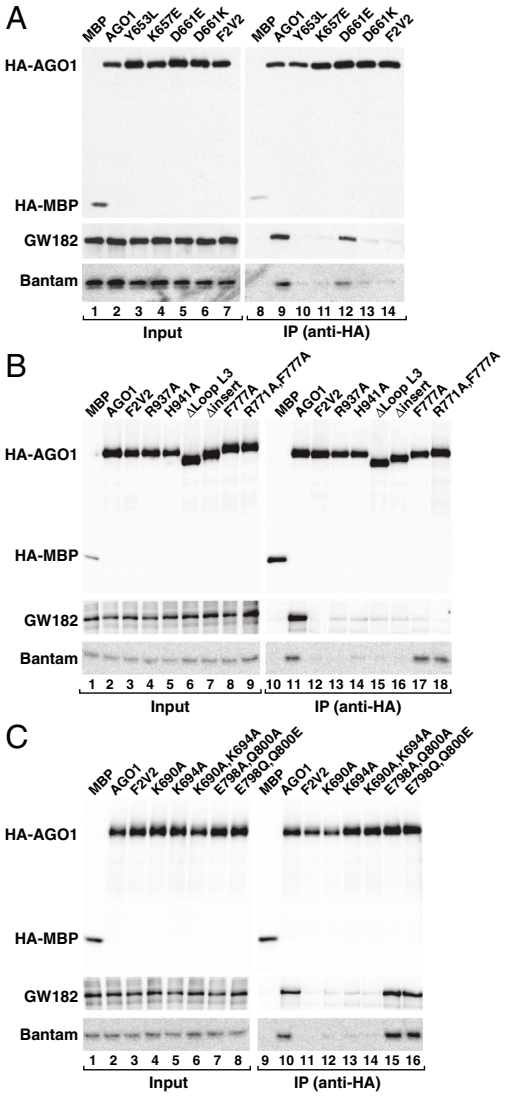


Fig. 5. Mutational analysis of *Dm* AGO1. (*A*–*C*) Lysates from S2 cells expressing HA-tagged versions of MBP, wild-type AGO1 or AGO1 mutants were immunoprecipitated using a monoclonal anti-HA antibody. Inputs and immunoprecipitates were analyzed by Western blotting. Endogenous GW182 was detected using anti-GW182 antibodies. The association between HA-AGO1 and endogenous Bantam miRNA was analyzed by northern blotting.

Strikingly, we could not identify any mutation that disrupts binding to miRNAs without simultaneously affecting GW182 binding (Fig. 5 *A*–*C*). This correlation is strict, but the converse is not the case; there are mutations that reduce GW182 binding whereas miRNA binding remains intact [R771A (K720) and F777A (W726), as reported previously (Fig. 5*B*, lanes 17 and 18; Figs. S1 and S5 and ref. 27]. Consequently, miRNA binding may be required for GW182 binding, whereas GW182 binding might facilitate, but is not required, for miRNA binding. Notably, these results also exclude the possibility that the loss of miRNA binding after mutating D661 (D603) or any other interface residue indirectly results from a loss of GW182 binding. Clearly, an understanding of the complex interplay between miRNA binding and the recruitment of GW182 requires detailed structural information.

Conclusion

The structure of a MID-PIWI lobe from a eukaryotic AGO illustrates that the relative domain orientations are highly conserved compared with the prokaryotic structures, indicating a conserved mode of interaction with the RNA substrate. The interface of the two domains is polar and hydrophilic, with D603 as a central residue that is crucial for the stability of the MID-PIWI interaction. This argues against the direct involvement of the analogous residue in the allosteric regulation of human AGO2 or *Dm* AGO1 and provides a simple explanation for why the D661K (D603K) mutation abolished the small RNA binding activity of *Dm* AGO1 in vivo. Interestingly, the more subtle H941A (H899) mutation had a similar, although slightly weaker effect, confirming that the MID-PIWI interface is conserved and quite sensitive to alterations. Mutational analysis of the *Dm* AGO1 protein also showed that GW182 binding is lost in all instances in which miRNA binding is lost, indicating that GW182 binding depends on the presence of miRNA or an AGO conformation that results from the presence of miRNA.

In conclusion, the hydrophilic MID-PIWI interface of *Nc* QDE2 (comprising residues R895-D603-H899) is particularly well conserved in AGOs from the animal miRNA pathway, and its stability has direct consequences for substrate binding. This raises the possibility that factors that influence the stability of the interface might have evolved to regulate AGO function. The interface itself may also be dynamic during the AGO reaction cycle. This would fit with a flexible role for the AGO proteins, where the adoption of guide and/or target specific conformations (30–32) would channel them into distinct pathways (such as the miRNA or siRNA pathway).

Materials and Methods

Detailed experimental procedures are given in *SI Text*. Briefly, the MID-PIWI lobe of *Nc* QDE-2 (amino acids 506–938) was expressed from a pETM60 vector (derived from pET24d; Novagen) in *Escherichia coli* BL21 Star (DE3) cells as a NusA-6xHis-tagged fusion protein. It was purified by a Ni^{2+}-affinity step (HiTrap Chelating HP column, GE Healthcare), followed by a removal of the purification tag and additional cation exchange (SP) chromatography and gel filtration steps (HiLoad 26/60 Superdex 75 pg; GE Healthcare). The proteins was concentrated to 20 mg/mL in 10 mM Hepes (pH 7.2), 150 mM NaCl and 1 mM DTT.

The structure of the MID-PIWI lobe was solved by molecular replacement using the MID domain (PDB ID code 2xdy; ref. 17) as a search model. The resulting difference density allowed the placement of a copy of the *T. thermophilus* PIWI domain [from PDB ID code 3DLB (8)] to start model building and refinement. The structure of the MID-PIWI ΔL3 was solved by molecular replacement using the low-resolution structure as a model. For analytical size exclusion chromatography, proteins or mixtures of proteins with RNA were injected onto the columns and UV absorption was detected simultaneously at 230, 260, and 280 nm.

Mutants of *D. melanogaster* AGO1 were generated by site-directed mutagenesis using the plasmid pAc5.1B-λN-HA-AGO1 as template and the QuickChange Mutagenesis Kit from Stratagene. All *Dm* AGO1 mutants and the equivalent QDE-2 mutants are described in Table S2. The interaction of AGO1 with endogenous miRNAs and GW182 as well as the complementation assays were performed as described previously (27, 29).

ACKNOWLEDGMENTS. We thank G. Macino for providing *N. crassa* QDE-2 cDNA and R. Büttner for excellent technical assistance. We thank the staff at the PX beamlines of the Swiss Light Source for assistance with data collection. This study was supported by the Max Planck Society, by grants from the Deutsche Forschungsgemeinschaft (DFG, FOR855 and the Gottfried Wilhelm Leibniz Program awarded to E.I.) and by the Sixth Framework Programme of the European Commission, through the SIROCCO Integrated Project LSHG-CT-2006-037900.

1. Tolia NH, Joshua-Tor L (2007) Slicer and the argonautes. *Nat Chem Biol* 3:36–43.
2. Jínek M, Doudna JA (2009) A three-dimensional view of the molecular machinery of RNA interference. *Nature* 457:405–412.
3. Bartel PD (2009) MicroRNAs: Target recognition and regulatory functions. *Cell* 136:215–233.
4. Song JJ, Smith SK, Hannon GJ, Joshua-Tor L (2004) Crystal structure of Argonaute and its implications for RISC slicer activity. *Science* 305:1434–1437.
5. Ma JB, et al. (2005) Structural basis for 5′-end-specific recognition of guide RNA by the A. fulgidus Piwi protein. *Nature* 434:666–670.
6. Parker JS, Roe SM, Barford D (2005) Structural insights into mRNA recognition from a PIWI domain-siRNA guide complex. *Nature* 434:663–666.
7. Yuan YR, et al. (2005) Crystal structure of A. aeolicus argonaute, a site-specific DNA-guided endoribonuclease, provides insights into RISC-mediated mRNA cleavage. *Mol Cell* 19:405–419.
8. Wang Y, Sheng G, Juranek S, Tuschl T, Patel DJ (2008) Structure of the guide-strand-containing argonaute silencing complex. *Nature* 456:209–213.
9. Wang Y, et al. (2008) Structure of an argonaute silencing complex with a seed-containing guide DNA and target RNA duplex. *Nature* 456:921–926.
10. Wang Y, et al. (2009) Nucleation, propagation and cleavage of target RNAs in Ago silencing complexes. *Nature* 461:754–761.
11. Parker JS, Roe SM, Barford D (2004) Crystal structure of a PIWI protein suggests mechanisms for siRNA recognition and slicer activity. *EMBO J* 23:4727–4737.
12. Lingel A, Simon B, Izaurralde E, Sattler M (2003) Structure and nucleic-acid binding of the *Drosophila* Argonaute 2 PAZ domain. *Nature* 426:465–469.
13. Lingel A, Simon B, Izaurralde E, Sattler M (2004) Nucleic acid 3′-end recognition by the Argonaute2 PAZ domain. *Nat Struct Mol Biol* 11:576–577.
14. Song JJ, et al. (2003) The crystal structure of the Argonaute2 PAZ domain reveals an RNA binding motif in RNAi effector complexes. *Nat Struct Mol Biol* 10:1026–1032.
15. Yan KS, et al. (2003) Structure and conserved RNA binding of the PAZ domain. *Nature* 426:468–474.
16. Ma JB, Ye K, Patel DJ (2004) Structural basis for overhang specific small interfering RNA recognition by the PAZ domain. *Nature* 429:318–322.
17. Boland A, Tritschler F, Heimstädt S, Izaurralde E, Weichenrieder O (2010) Crystal structure and ligand binding of the MID domain of a eukaryotic Argonaute protein. *EMBO Rep* 11:522–527.
18. Frank F, Sonenberg N, Nagar B (2010) Structural basis for 5′-nucleotide base-specific recognition of guide RNA by human AGO2. *Nature* 465:818–822.
19. Simon B, et al. (2011) Recognition of 2′-O-methylated 3′-end of piRNA by the PAZ domain of a Piwi protein. *Structure* 19:172–180.
20. Tian Y, Simanshu DK, Ma JB, Patel DJ (2011) Structural basis for piRNA 2′-O-methylated 3′-end recognition by Piwi PAZ(Piwi/Argonaute/Zwille) domains. *Proc Natl Acad Sci USA* 108:903–910.
21. Djuranovic S, et al. (2010) Allosteric regulation of Argonaute proteins by miRNAs. *Nat Struct Mol Biol* 17:144–150.
22. Fulci V, Macino G (2007) Quelling: post-transcriptional gene silencing guided by small RNAs in Neurospora crassa. *Curr Opin Microbiol* 10:199–203.
23. Lee HC, et al. (2010) Diverse pathways generate microRNA-like RNAs and Dicer-independent small interfering RNAs in fungi. *Mol Cell* 38:803–814.
24. Parker JS, Parizotto EA, Wang M, Roe SM, Barford D (2009) Enhancement of the seed-target recognition step in RNA silencing by a PIWI/MID domain protein. *Mol Cell* 33:204–214.
25. Lambert NJ, Gu SG, Zahler AM (2011) The conformation of microRNA seed regions in native microRNPs is prearranged for presentation to mRNA targets. *Nucleic Acids Res* 10.1093/nar/gkr077.
26. Huntzinger E, Izaurralde E (2011) Gene silencing by microRNAs: Contributions of translational repression and mRNA decay. *Nat Rev Genet* 12:99–110.
27. Eulalio A, Helms S, Fritzsch C, Fauser M, Izaurralde E (2009) A C-terminal silencing domain in GW182 is essential for miRNA function. *RNA* 15:1067–1077.
28. Till S, et al. (2007) A conserved motif in Argonaute-interacting proteins mediates functional interactions through the Argonaute PIWI domain. *Nat Struct Mol Biol* 14:897–903.
29. Eulalio A, Huntzinger E, Izaurralde E (2008) GW182 interaction with Argonaute is essential for miRNA-mediated translational repression and mRNA decay. *Nat Struct Mol Biol* 15:346–353.
30. Chen HM, et al. (2010) 22-Nucleotide RNAs trigger secondary siRNA biogenesis in plants. *Proc Natl Acad Sci USA* 107:15269–15274.
31. Cuperus JT, et al. (2010) Unique functionality of 22-nt miRNAs in triggering RDR6-dependent siRNA biogenesis from target transcripts in Arabidopsis. *Nat Struct Mol Biol* 997–1003.
32. Noto T, et al. (2010) The Tetrahymena argonaute-binding protein Giw1p directs a mature argonaute-siRNA complex to the nucleus. *Cell* 140:692–703.

Supporting Information

Boland et al. 10.1073/pnas.1103946108

SI Materials and Methods

Cloning, Expression, and Purification of the QDE-2 MID-PIWI Lobe. A cDNA fragment encoding the MID and PIWI domains of *Nc* QDE-2 (amino acids 506–938) was amplified by PCR from a pBluescript-SK plasmid containing the genomic sequence of the *qde-2* gene (kindly provided by G. Macino) using the primer pairs:

5′ATGCCATGGCAGTCAAGGTCGCCAGACCTT-3′ (forward) and 5′- TAGCGGCCGCTTAGATATAGTACATGGAGTTCCTA -3′ (reverse). The cDNA was then cloned into the NcoI and NotI sites of the pETM60 vector (derived from pET24d; Novagen), downstream of a NusA-6xHis tag. The construct for the expression of the QDE-2 MID-PIWI ΔL3 protein in which residues 787–838 were replaced by a Gly-Ser-Gly linker was obtained by site-directed mutagenesis.

The NusA-6xHis-tagged MID-PIWI lobe proteins were expressed in *Escherichia coli* BL21 Star (DE3) cells (Invitrogen) at 20 °C overnight. Protein expression was induced with 0.5 mM Isopropyl ß-D-1-thiogalactopyranoside at an A_{600} of 0.6. The fusion proteins were first purified by a Ni^{2+}-affinity step (HiTrap Chelating HP column, GE Healthcare). Removal of the NusA and polyhistidine tags was done by proteolytic digestion with tobacco etch virus protease at 4 °C overnight. The proteins were further purified by cation exchange (SP) chromatography and subsequent gel filtration (HiLoad 26/60 Superdex 75 pg; GE Healthcare). Finally, the proteins were concentrated to 20 mg/mL in 10 mM Hepes (pH 7.2), 150 mM NaCl and 1 mM DTT.

Crystallization. Crystals of native and selenomethionine-substituted QDE-2 MID-PIWI construct were grown by hanging drop vapor diffusion over a 500 μL reservoir at 18 °C. The protein solution described above was mixed 1 : 1 with various reservoir solutions containing 1.6–2.3 M ammonium sulfate as precipitant and 100–250 mM potassium format leading to a total droplet size of 3 μL. Crystals were optimized by hair-seeding and flash frozen in liquid nitrogen in the respective reservoir solution supplemented with 15 or 20% glycerol for cryoprotection.

Diffraction quality crystals of the QDE-2 MID-PIWI ΔL3 protein were grown by sitting drop vapor diffusion over a 50 μL reservoir at 22 °C after mixing 0.2 μL protein solution with 0.2 μL reservoir solution (2.3 M ammonium sulphate and 200 mM di-sodium tartrate). Crystals were cryoprotected in 2.3 M ammonium sulfate and 200 mM di-sodium tartrate supplemented with 20% glycerol before flash-freezing in liquid nitrogen.

Data Collection, Structure Solution, and Refinement. Diffraction data were recorded on a PILATUS 6M detector (1) at the beamline PXII of the Swiss Light Source at a wavelength of 1.0 Å for the native dataset and at the absorption peak of the selenomethionine derivative at 0.9792 Å. Extensive screening of crystals from the QDE-2 MID-PIWI protein led to a native dataset at 3.65 Å resolution that was assembled from three 10° wedges collected from different spots along the crystal due to severe radiation damage. The best selenomethionine dataset extended to 4.35 Å resolution. Diffraction images were processed with XDS (2).

The structure was solved by molecular replacement in PHASER (3), using the MID domain [Protein Data Bank (PDB) ID code 2xdy (4)] as a search model. A copy of the *Thermus thermophilus* PIWI domain [from PDB ID code 3DLB (5)] could be placed in the resulting difference density and was rigid-body refined and trimmed in COOT (6). Aided by the high solvent content (79%) of this crystal form, we could build (COOT) and refine a model of QDE-2 MID-PIWI using PHENIX (7). Because loop L3 was disordered in this crystal form, we deleted the relevant sequence in the QDE-2 MID-PIWI ΔL3 construct. The resulting crystals diffracted X-rays to 1.85 Å resolution, and the structure was solved by molecular replacement (PHASER) using the low-resolution structure as a model. The QDE-2 MID-PIWI ΔL3 model was (re-)built automatically with ARP/wARP (8) and finished manually in COOT, alternating with rounds of refinement (PHENIX). Finally, the high-resolution QDE-2 MID-PIWI ΔL3 model was used for a last cycle of model adjustment and refinement of the low-resolution QDE-2 MID-PIWI model, and proper sequence assignment was confirmed by the calculation of anomalous difference Fourier maps from the selenomethionine dataset, with peaks corresponding to the methionine positions. Stereochemical properties were analyzed with MOLPROBITY (9) and WHATCHECK (10), and structure figures were generated in PYMOL.

Analytical Size Exclusion Chromatography. Proteins were prepared and analyzed in the same buffer, as described for crystallization (10 mM Hepes (pH 7.2), 150 mM NaCl, 1 mM DTT). Proteins or mixtures of proteins with RNA were injected either onto a Superdex 75 10/300 (MID domain) or onto a Superdex 200 10/300 GL (MID-PIWI) analytical gel filtration column (GE Healthcare, 18 °C, preequilibrated) as part of an ÄKTA Purifier-10 at a flow rate of 0.5 mL/min. UV absorption was detected simultaneously at 230 nm, 260 nm and 280 nm. Protein and nucleic acid concentrations were estimated from the theoretical molar extinction coefficients at 280 nm and 260 nm, respectively. The relative contributions of protein and nucleic acid to the total absorption at each wavelength were calculated assuming constant ratios of the extinction coefficient at 230 nm to the extinction coefficient at 280 nm for each substance (11). The protein mutants tested are described in Table S2.

Coimmunoprecipitation and Complementation Assays in S2 Cells. Mutants of *Drosophila melanogaster* AGO1 were designed based on the structure of the *Nc* QDE-2 MID-PIWI lobe with the goal to avoid general folding defects. Mutants were generated by site-directed mutagenesis using the plasmid pAc5.1B-λN-HA-AGO1 small interfering RNA (siRNA)-resistant as template and the QuickChange Mutagenesis Kit from Stratagene. All *Dm* AGO1 mutants and the equivalent QDE-2 mutants are described in Table S2. Transfections of S2 cells were performed in 6-well plates, using Effectene transfection reagent (Qiagen). The interaction of AGO1 with endogenous micro-RNAs (miRNAs) and GW182 was tested as described previously (12). Wild-type AGO1 or AGO1 mutants were immunoprecipitated using a monoclonal antiHA antibody (Covance Research Products; catalog number MMS-101P). Inputs (1.5%) and immunoprecipitates (30%) were analyzed by Western blotting using a polyclonal antiHA antibody (SIGMA; catalog number T6199). Antibodies to *D. melanogaster* GW182 were described before (13). All Western blot experiments were developed with the ECL Western Blotting Detection System (GE Healthcare) as recommended by the manufacturer.

Complementation assays were carried out as described before (12, 14) except that transfections were performed in 24-well plates. The following siRNAs were used: control GFP siRNA (5′-GCGACGUAAACGGCCACAAGUUCUU) and AGO1 siRNA1 (5′- CGAAGGAGAUCAAGGGUUUUU) at a final concentration of 75 nM. The transfection mixtures contained 20 ng of F-Luc-Par-6 reporter plasmid, 80 ng of the *Renilla* transfection

control, and 20 ng of a plasmid expressing miR-1 primary transcript or the corresponding vector without insert. When indicated, 10 ng of plasmid expressing siRNA-resistant forms of wild-type or mutant *Dm* AGO1 were included. A plasmid expressing Maltose binding protein (MBP) served as a negative control.

1. Broennimann C, et al. (2006) The PILATUS 1M detector. *J Synchrotron Radiat* 13:120–130.
2. Kabsch W (2010) XDS. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):125–132.
3. McCoy AJ, et al. (2007) Phaser crystallographic software. *J Appl Crystallogr* 40(Pt 4):658–674.
4. Boland A, Tritschler F, Heimstädt S, Izaurralde E, Weichenrieder O (2010) Crystal structure and ligand binding of the MID domain of a eukaryotic Argonaute protein. *EMBO Rep* 11:522–527.
5. Wang Y, Sheng G, Juranek S, Tuschl T, Patel DJ (2008) Structure of the guide-strand containing argonaute silencing complex. *Nature* 456:2093–213.
6. Emsley P, Cowtan K (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60:2126–2132.
7. Adams PD, et al. (2010) PHENIX: A comprehensive Phython-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):213–221.
8. Cohen SX, et al. (2008) ARP/wARP and molecular replacement: The next generation. *Acta Crystallogr D Biol Crystallogr* 64:49–60.
9. Davis IW, et al. (2007) MolProbity: All-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* 35:W375–W383.
10. Hooft RW, Vriend G, Sander C, Abola EE (1996) Errors in protein structures. *Nature* 381:272.
11. Müller M, Weigand JE, Weichenrieder O, Suess B (2006) Thermodynamic characterization of an engineered tetracycline-binding riboswitch. *Nucleic Acids Res* 34:2607–2617.
12. Eulalio A, Huntzinger E, Izaurralde E (2008) GW182 interaction with Argonaute is essential for miRNA-mediated translational repression and mRNA decay. *Nat Struct Mol Biol* 15:346–353.
13. Behm-Ansmant I, et al. (2006) mRNA degradation by miRNAs and GW182 requires both CCR4:NOT deadenylase and DCP1:DCP2 decapping complexes. *Genes Dev* 20:1885–1898
14. Eulalio A, Helms S, Fritzsch C, Fauser M, Izaurralde I (2009) A C-terminal silencing domain in GW182 is essential for miRNA function. *RNA* 15:1067–1077.

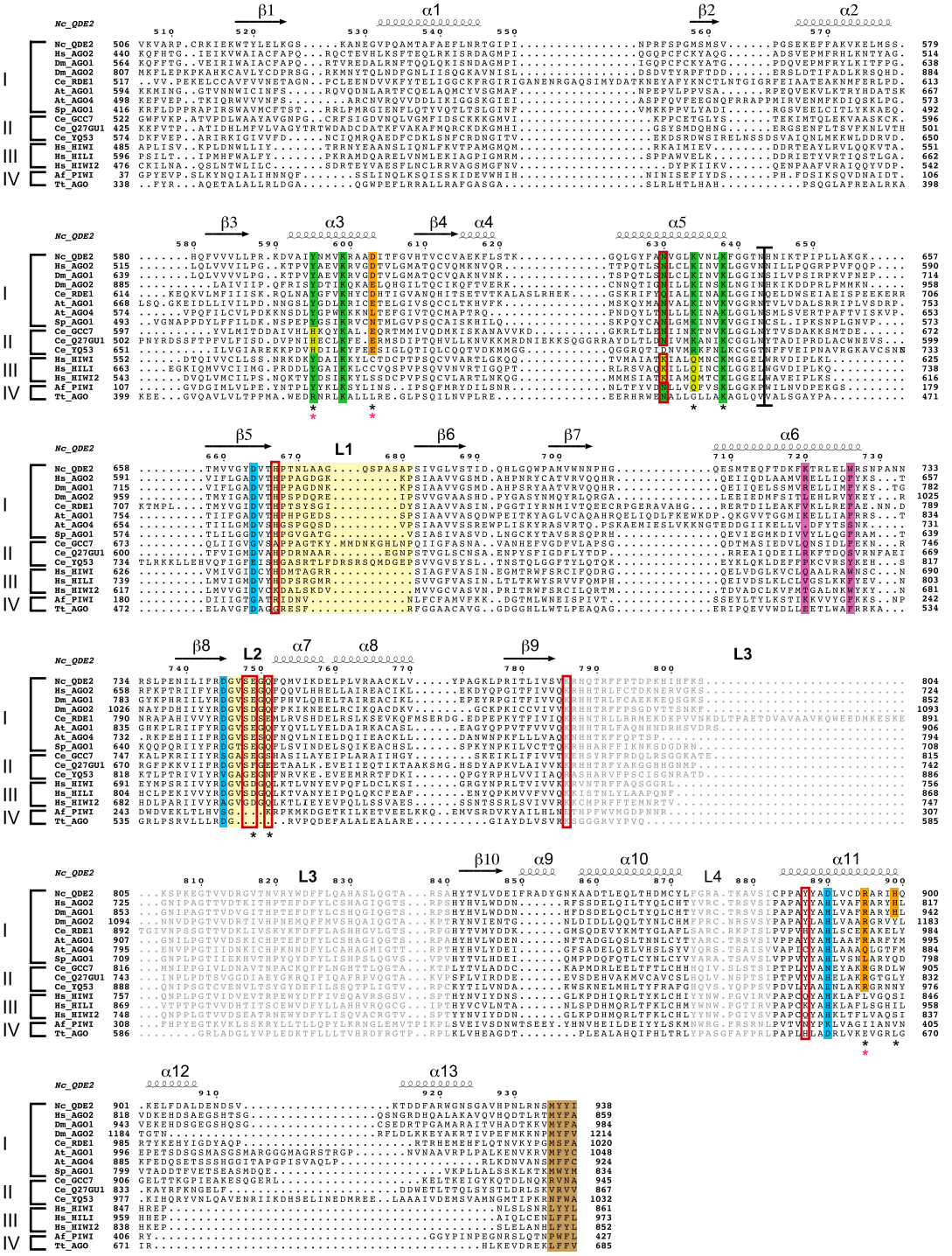


Fig. S1. Structure-based sequence alignment of the *Nc* QDE-2 MID and PIWI domains with archaeal and eubacterial homologs (IV), as well as with eukaryotic MID-PIWI domains of AGOs from the Argonaute-like clade (I), the WAGO clade (II), and the PIWI-like clade (III). The MID-PIWI domain boundary is indicated with a vertical black line. Residues lining the 5′ binding pocket are shaded green, and residues involved in interdomain interactions are shaded orange. Clade-specific differences are highlighted in pale green. Loop L1 and L2 are shaded yellow. Residues in the unstructured loops L3 and L4 are indicated in gray. The catalytic residues (DDD) are shaded cyan. The C-terminal conserved hydrophobic residues are shaded brown. The eukaryotic insertion includes α-helices 12 and 13. Mutations that abolish GW182 binding but not miRNA binding (14), are shaded magenta. Residues potentially involved in RNA binding are boxed in red. Residues mutated in the QDE-2 MID-PIWI domain are marked with a red asterisk, and residues mutated in *Dm* AGO1 are labeled with a black asterisk. AGO, Argonaute; *Af, Archaeoglobus fulgidus; At, Arabidopsis thaliana; Ce, Caenorhabditis elegans; Dm, Drosophila melanogaster; Hs, Homo sapiens; Nc, Neurospora crassa*; PIWI, P-element-induced wimpy testis; RDE-1, RNA interference deficient 1; *QDE-2, quelling-defective 2; Sp, Schizosaccharomyces pombe; Tt, Thermus thermophilus.*

UniProt accession numbers: *Nc* QDE2 (Q9P8T1), *Hs* AGO2 (Q9UKV8), *Dm* AGO1 (Q32KD4), *Dm* AGO2 (Q9VUQ5), *Ce* RDE1 (Q9XU82), *At* AGO1 (O04379), *At* AGO4 (Q9ZVD5), *Sp* AGO1 (O74957), *Ce* GCC7 (Q21770), *Ce* Q27GU1 (Q27GU1), *Ce* YQ53 (Q09249), *Hs* HIWI (Q96J94), *Hs* HILI (Q8TC59), *Hs* HIWI2 (Q7Z3Z4), *Af* PIWI (O28951), *Tt* AGO (Q746M7).

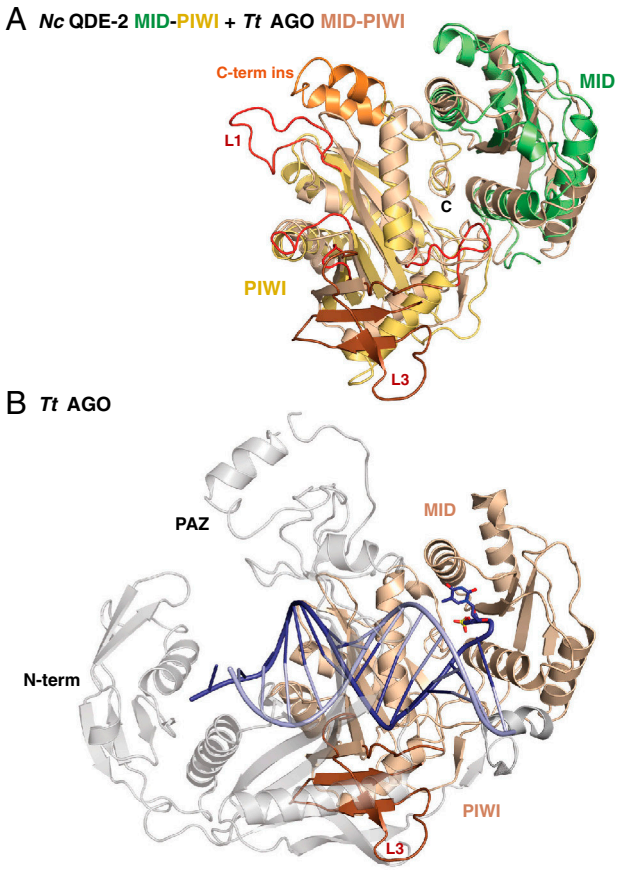


Fig. S2. (*A*) Structural superposition of a *Tt* AGO-nucleic acid complex (as shown in Fig. 1*C*) onto the QDE-2 MID-PIWI lobe (using the PIWI domains as reference). The guide DNA-target RNA duplex was removed. (*B*) Ribbon representation of the *Tt* AGO-nucleic acid complex. The N-terminal and PAZ domains are shown in gray, and the MID-PIWI lobe in brown. Loop L3, which is highlighted in dark brown, contacts the N-terminal domain as well as the PAZ-MID domain linker. Guide strand: dark blue; target strand: light blue [PDB ID code 3HJF (1)].

1 Wang Y, et al. (2009) Nucleation, propagation, and cleavage of target RNAs in Ago silencing complexes. *Nature* 461:754–761.

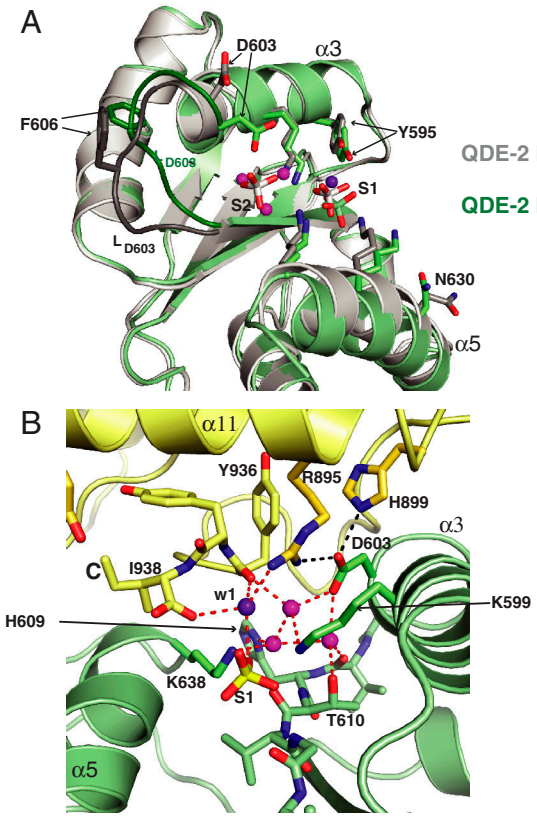


Fig. S3. Structural superposition of the isolated QDE-2 MID domain (gray) onto the QDE-2 MID-PIWI lobe (green). (*A*) Loop L_{D603} changes conformation between the two structures. The structure of the isolated QDE-2 MID domain contains two bound sulfate ions (S1 and S2). S1 marks the position of the 5′ phosphate of the guide strand and is also present in the structure of the QDE-2 MID-PIWI lobe. S2 marks the position of a putative second ligand-binding site in the isolated MID domain [PDB ID code 2xdy (4)]. In the structure of the MID-PIWI lobe, three well-coordinated water molecules (magenta spheres) replace S2 and superimpose almost perfectly with three oxygens of S2. (*B*) Extended view of residues coordinating the water molecules including water w1.

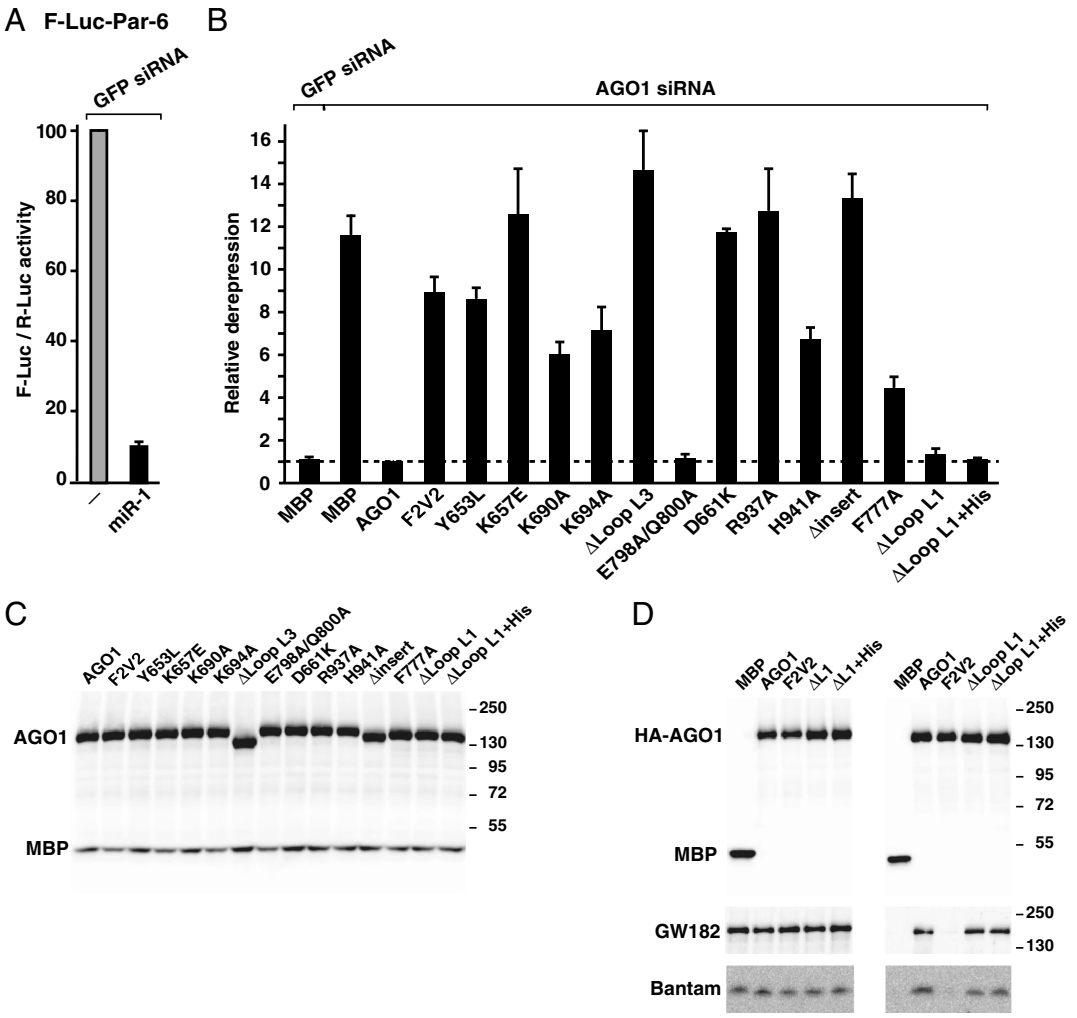


Fig. S4. Complementation assay in S2 cells. (*A–C*) S2 cells were transfected with the indicated siRNAs plus a mixture of three plasmids: one expressing the FLuc-Par-6 reporter; another expressing miR-1 primary transcript (black bars) or the corresponding empty vector (gray bar); and a third expressing *Renilla* luciferase (RLuc). Plasmids encoding siRNA-resistant versions of wild-type or mutant HA-AGO1 proteins were included in the transfection mixtures, as indicated. HA-MBP served as a negative control. Firefly luciferase activities were normalized to those of the *Renilla* luciferase transfection control and set to 100 in cells transfected with the empty vector (i.e., in the absence of the miR-1). Panel (*A*) shows normalized Firefly luciferase activities in the absence or presence of miR-1 in control cells (i.e., cells treated with GFP siRNA and transfected with HA-MBP). Panel (*B*) shows the relative fold derepression for each condition. Mean values ± standard deviations from three independent experiments are shown. (*C*) Western blot showing that all proteins were expressed at similar levels. (*D*) Lysates from S2 cells expressing HA-tagged versions of MBP, wild-type AGO1 or the indicated AGO1 mutants were immunoprecipitated using a monoclonal antiHA antibody and analyzed as described in Fig. 5 *A–C*.

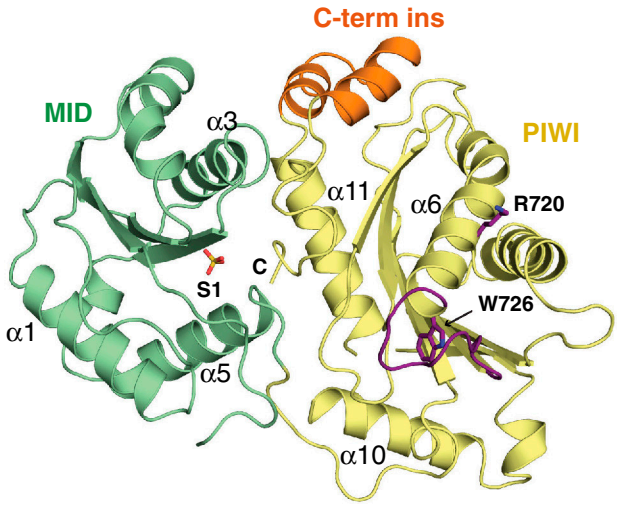


Fig. S5. Dm AGO1 residues that impair GW182 binding without compromising miRNA binding. Ribbon representation of the *Nc* QDE-2 MID-PIWI lobe rotated 180° relative to the view in Fig. 1. The MID and PIWI domains are shown in green and yellow respectively. The C-terminal insertion is shown in orange. Elements that are important for GW-binding but not for microRNA-binding are displayed in purple.

Table S1. Data collection and refinement statistics

Data set	MID-PIWI	MID-PIWI ΔLOOP
	(Crystal form I)	(Crystal form II)
Data collection*		
Space group	$P6_322$	$P3_221$
Unit cell		
dimensions (a/b/c), Å	218.9/218.9/76.7	63.1/63.1/170.5
angles (α/β/γ), °	90/90/120	90/90/120
Resolution range, Å	54.7–3.65 (3.75–3.65)	54.6–1.85 (1.94–1.85)
R_{sym}, %	11.0 (53.6)	5.6 (66.5)
Completeness, %	96.1 (97.1)	99.8 (100)
Mean $I/\sigma(I)$	9.1 (2.2)	15.3 (2.4)
Number of unique reflections	12,017	34,431
Multiplicity	3.2	5.1
Refinement		
Data range, Å	54.7–3.65	54.6–1.85
R_{cryst}, %	23.3	19.6
R_{free}, %	25.2	23.6
Number of atoms		
per asymmetric unit		
all atoms	2,891	3,066
protein	2,880	2,841
ligand/ion	6/5	6/5
water	0	225
Average *B*-factor, Å^2		
all atoms	99.8	36.8
protein	99.8	36.5
ligand/ion	107.5	48.2
water	—	40.0
Ramachandran plot		
favored regions, %	97.0	97.5
disallowed regions, %	0.0	0.0
Rmsd from ideal geometry		
bond lengths, Å	0.015	0.015
bond angles, °	1.24	1.45

*Values in parentheses are for highest-resolution shell.

Table S2. Overview of *Dm* AGO1 mutants used in this study and equivalent residues in QDE-2

Dm AGO1	*Nc* QDE-2 equivalent residues	Predicted function
Y653L	Y595L	5′ binding pocket / guide RNA binding
K657E	K599	5′ binding pocket / guide RNA binding
K690A	K634	5′ binding pocket / guide RNA binding
K694A	K638	5′ binding pocket / guide RNA binding
R937A	R895A	MID-PIWI interface and 5′ binding pocket
D661ED661K	D603	MID-PIWI interface
H941AH941QH941K	H899A	MID-PIWI interface
E798A, Q800AE798Q, Q800E	E749, Q751	target strand binding
ΔL3 (835–886)	ΔL3 (787–838)	RNA binding, contacts the N-term-PAZ lobe
ΔC-term insertion(942–971)	ΔC-term insertion(900–925)	predicted to contact the N-term-PAZ lobe
ΔL1 (726–733)	ΔL1 (669–681)	predicted to contact the N-term-PAZ lobe
ΔL1+His724 (724–733)	ΔL1+His667 (667–681)	predicted to contact the N-term-PAZ lobe and the guide-target RNA duplex

Molecular Cell
Article

Structure of the PAN3 Pseudokinase Reveals the Basis for Interactions with the PAN2 Deadenylase and the GW182 Proteins

Mary Christie,[1,2] Andreas Boland,[1,2] Eric Huntzinger,[1] Oliver Weichenrieder,[1,*] and Elisa Izaurralde[1,*]
[1]Department of Biochemistry, Max Planck Institute for Developmental Biology, Spemannstrasse 35, 72076 Tübingen, Germany
[2]These authors contributed equally to this work
*Correspondence: oliver.weichenrieder@tuebingen.mpg.de (O.W.), elisa.izaurralde@tuebingen.mpg.de (E.I.)
http://dx.doi.org/10.1016/j.molcel.2013.07.011

SUMMARY

The PAN2-PAN3 deadenylase complex functions in general and miRNA-mediated mRNA degradation and is specifically recruited to miRNA targets by GW182/TNRC6 proteins. We describe the PAN3 adaptor protein crystal structure that, unexpectedly, forms intertwined and asymmetric homodimers. Dimerization is mediated by a coiled coil that links an N-terminal pseudokinase to a C-terminal knob domain. The PAN3 pseudokinase binds ATP, and this function is required for mRNA degradation in vivo. We further identified conserved surfaces required for mRNA degradation, including the binding surface for the PAN2 deadenylase on the knob domain. The most remarkable structural feature is the presence of a tryptophan-binding pocket at the dimer interface, which mediates binding to TNRC6C in human cells. Together, our data reveal the structural basis for the interaction of PAN3 with PAN2 and the recruitment of the PAN2-PAN3 complex to miRNA targets by TNRC6 proteins.

INTRODUCTION

Shortening of mRNA poly(A) tails by deadenylation is a common mechanism for regulating the translation and stability of mRNAs (Weill et al., 2012). Deadenylation represses translation and can initiate irreversible mRNA degradation. Indeed, eukaryotic bulk mRNA degradation begins with the removal of the mRNA poly(A) tail; deadenylated mRNAs are then decapped and degraded by the 5′-to-3′ exonuclease XRN1. Alternatively, following deadenylation mRNAs can be degraded from the 3′ end by the exosome (Wahle and Winkler, 2013).

Current models suggest that mRNAs are deadenylated by the consecutive but redundant action of two cytoplasmic deadenylase complexes: PAN2-PAN3 and CCR4-NOT (Tucker et al., 2001; Yamashita et al., 2005). Poly(A) tails are first trimmed to approximately 50–110 nt by the distributive activity of the PAN2-PAN3 complex, after which the shortened mRNA tails are rapidly degraded by the CCR4-NOT complex (Tucker et al., 2001; Yamashita et al., 2005). Although the PAN2-PAN3 complex can associate with the CCR4-NOT complex (Zheng et al., 2008), the molecular details and mechanism for the consecutive action of these complexes are unclear.

The PAN2-PAN3 and CCR4-NOT complexes commonly feature catalytically active deadenylases in association with noncatalytic subunits that are required for efficient deadenylation in vivo and likely confer additional regulatory functions. However, the precise role of the noncatalytic subunits in mRNA deadenylation is not completely understood. One possible function is the stabilization of the catalytic subunits (Wahle and Winkler, 2013). A second possibility is the recruitment of the catalytic subunits to specific mRNA targets (Wahle and Winkler, 2013; Weill et al., 2012). One example is the N-terminal PAM2 motif (PABP-interacting motif 2) of PAN3, which interacts with the cytoplasmic poly(A) binding protein (PABP) bound to the poly(A) tails of mRNAs. This interaction enables PAN3 to recruit the PAN2 deadenylase to polyadenylated mRNAs to initiate their deadenylation (Mangus et al., 2004; Siddiqui et al., 2007; Uchida et al., 2004). Accordingly, PAN2 is inactive in the absence of PAN3, and the activity of the PAN2-PAN3 complex is stimulated by PABP (Boeck et al., 1996; Brown et al., 1996; Lowell et al., 1992; Uchida et al., 2004; Zheng et al., 2008).

In addition to the interaction between PAN3 and PABP, which plays a role in general mRNA deadenylation, other examples of RNA-associated proteins that recruit the PAN2-PAN3 complex to specific mRNA targets to expedite their degradation have been described. For example, PAN3 cooperates with the Dun1 kinase to regulate the expression of *RAD5* mRNA (Hammet et al., 2002). Furthermore, proteins of the GW182 family (known as TNRC6A–C in humans), which are required for miRNA-mediated gene silencing in animals, interact directly with PAN3, thereby recruiting the PAN2-PAN3 complex to miRNA targets (Braun et al., 2011; Chekulaeva et al., 2011; Fabian et al., 2011).

Despite the importance of deadenylation in posttranscriptional mRNA regulation, structural information is available for only a few subunits or domains of the CCR4-NOT complex (Wahle and Winkler, 2013). However, no structural information has been obtained for the PAN2-PAN3 complex. To gain better insight into the molecular mechanism of mRNA deadenylation, we determined the crystal structures of the folded domains of the PAN3 protein from *Neurospora crassa* (*Nc*) and *Drosophila melanogaster* (*Dm*). The crystal structures reveal that PAN3 homodimerizes through an unexpected coiled coil (CC) that

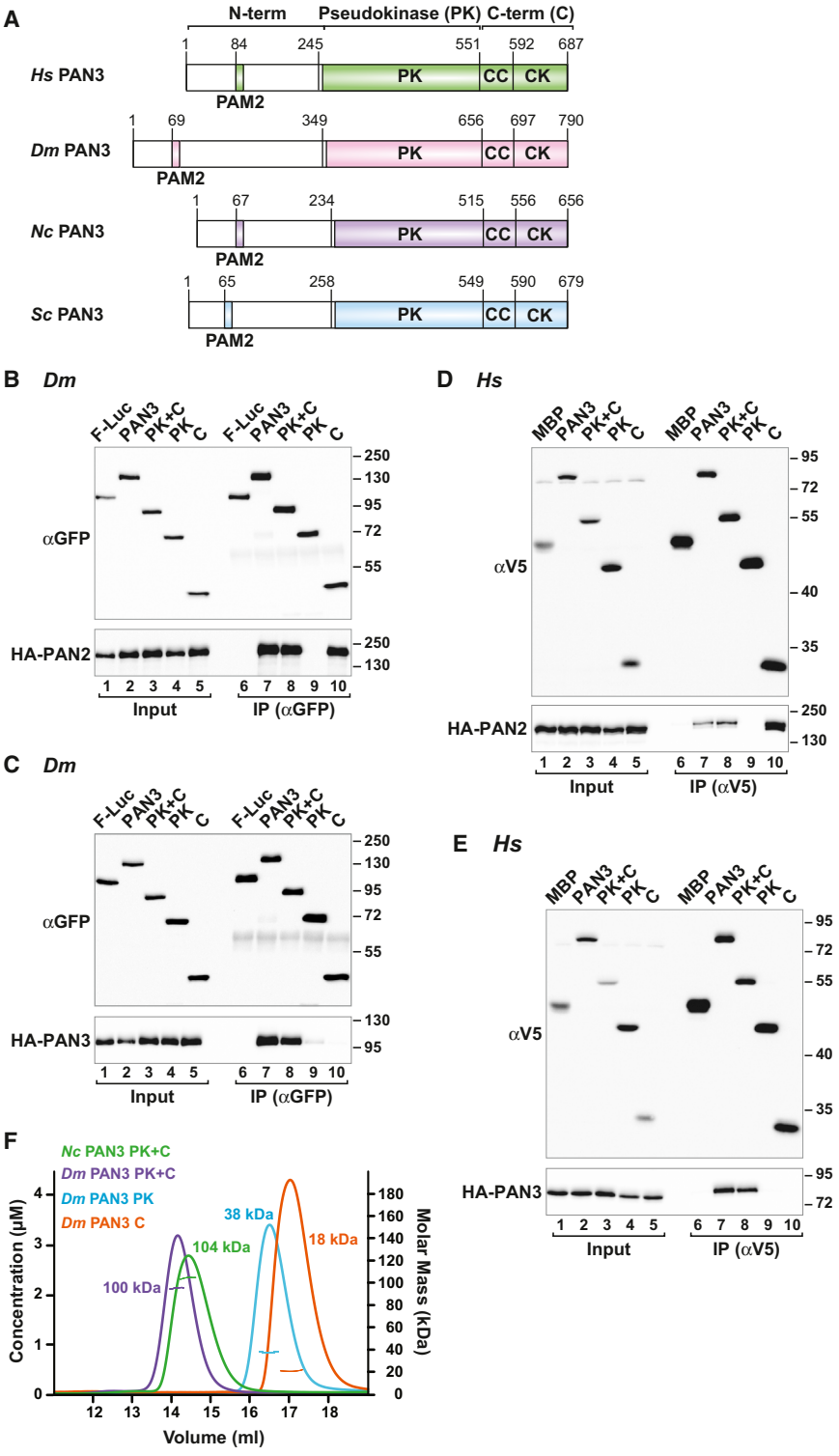


Figure 1. PAN3 Dimerizes and Interacts with PAN2

(A) Schematic representation of *Hs*, *Dm*, *Nc*, and *Sc* PAN3 showing the positions of the N-term region and the pseudokinase (PK) and C-term domains. The N-term contains a PAM2 motif. The C-term domain contains a coiled coil (CC) and a conserved C-terminal knob (CK) domain.

(B and C) Interaction between GFP-*Dm* PAN3 (full-length or fragments) and HA-*Dm* PAN2 or PAN3 in S2 cells. GFP-tagged firefly luciferase (F-Luc) served as a negative control.

(D and E) The interaction of V5-tagged *Hs* PAN3 (full-length or fragments) with HA-tagged *Hs* PAN2 and PAN3 proteins in human cells. V5-tagged maltose binding protein (MBP) served as a negative control.

(F) MALLS analysis of *Dm* PAN3 and *Nc* PAN3 protein fragments. The molecular weights of the proteins in solution are indicated. See also Table S1.

connects an N-terminal pseudokinase (PK) domain with a C-terminal knob domain (CK). Remarkably, the PAN3 PK domain has retained its ATP binding capacity, and this function is required for mRNA degradation in vivo. We used mutational and functional studies to identify key features of the *Dm* and *Hs* PAN3 protein. These studies revealed the PAN2 binding surface on PAN3, a W-binding pocket required for the interaction with TNRC6 proteins and additional conserved surface residues required for mRNA degradation. Collectively, our data provide a structural foundation for understanding the role of PAN3 in recruiting PAN2 to mRNA targets to initiate deadenylation.

RESULTS

PAN3 Dimerizes and Interacts with PAN2

PAN3 proteins contain three prominent regions: an unstructured N-terminal region (N-term), a central PK domain, and a highly conserved C-terminal domain (C-term) that is unique to the PAN3 protein family (Figure 1A). While the N-terminal region of PAN3 contains a well-characterized PAM2 motif (Mangus et al., 2004; Siddiqui et al., 2007), studies of the roles of the PK and C-term domains have primarily been restricted to the yeast protein. In particular, yeast two-hybrid assays have demonstrated that a C-terminal fragment of *Saccharomyces cerevisiae* (*Sc*) PAN3 (560–679) is required for *Sc* PAN2 binding, while both the PK and C-term domains mediate self-interaction (Mangus et al., 2004). However, the stoichiometry and significance of PAN3 oligomerization are still unknown.

To investigate whether these interactions are also conserved in *Dm* and human (*Hs*) PAN3 proteins, we performed coimmunoprecipitation assays using various PAN3 fragments in *Dm*

Table 1. Data Collection and Refinement Statistics

Data Set	*Dm* PAN3	*Nc* PAN3	*Nc* PAN3^{M}	*Nc* PAN3 (SeMet)
Space group	$P4_322$	$P6_5$	$P6_5$	$P6_5$
Unit cell				
Dimensions (a, b, c) (Å)	102.2, 102.2, 264.7	89.5, 89.5, 228.8	90.3, 90.3, 230.6	88.6, 88.6, 228.7
Angles (α , β , γ) (°)	90, 90, 90	90, 90, 120	90, 90, 120	90, 90, 120
Data Collection[a]				
Wavelength	0.99999	1.00001	0.99998	0.97927
Resolution range (Å)	49–3.6	46–3.3	46–2.85	46–4.0
R_{sym}, %	10.8 (89.1)	5.5 (76.8)	7.9 (72.6)	9.5 (78.6)
Completeness, %	99.8 (99.9)	99.9 (99.8)	99.9 (99.7)	99.9 (100)
Mean $I/\sigma(I)$	14.5 (2.1)	19.9 (3.0)	13.9 (2.2)	11.1 (2.7)
Unique reflections	17,015 (2,447)	15,611 (2,292)	24,793 (3,621)	8,626 (1,263)
Multiplicity	6.4 (6.6)	7.9 (7.7)	4.4 (4.4)	7.8 (7.6)
Refinement				
R_{cryst}, %	22.0	25.2	21.1	
R_{free}, %	26.9	29.4	24.2	
Number of atoms				
All atoms	6,125	6,407	6,524	
Protein	6,071	6,345	6,460	
Ligand	54	62	62/2	
Average B factor (Å^2)				
All atoms	136	119	89	
Protein	136	119	89	
Ligand	119	85	102/87	
Ramachandran plot				
Favored regions, %	96.1	96.6	96.4	
Disallowed regions, %	0.0	0.1	0.1	
Rmsd from ideal geometry				
Bond lengths (Å)	0.010	0.008	0.010	
Bond angles (°)	1.47	1.52	1.40	

[a]Values in parentheses are for highest-resolution shell.

Schneider (S2) and human HEK293 cells. We observed that a green fluorescent protein (GFP)-tagged *Dm* PAN3 fragment containing the PK and C-term domains (PK+C) displayed the binding properties of the full-length protein and interacted with HA (influenza hemagglutinin)-tagged *Dm* PAN2 and PAN3 (Figures 1B and 1C, lane 8 versus lane 7). Therefore, the PAN3 N-term is not required for these interactions. Further analyses demonstrated that the PAN3 C-term domain was sufficient for PAN2 binding (Figure 1B, lane 10), whereas both the PK and C-term domains were required for PAN3 self-interaction (Figure 1C, lane 8). Similarly, *Hs* PAN3 interacted with PAN2 via the C-term domain, which in isolation exhibited a higher affinity for PAN2 when compared to the PK+C fragment (Figure 1D, lane 10 versus lane 8). This differential binding was confirmed in independent experiments (data not shown) and suggests that the PK domain may negatively regulate the ability of the C-term domain to bind PAN2. By contrast, *Hs* PAN3 self-interaction required both the PK and C-term domains (Figure 1E, lane 8).

To determine the stoichiometry of PAN3 oligomers, we employed multiangle static laser-light scattering (MALLS) coupled with size-exclusion chromatography and measured the apparent molecular weight of the protein in solution. We found that the purified *Dm* PAN3 PK+C fragment was dimeric in solution (Figure 1F and see Table S1 online; molecular mass of 100 kDa compared to the expected value of 51 kDa for the monomeric protein). The corresponding *Nc* and *Hs* PAN3 fragments were also dimeric in solution (Figure 1F and Table S1). By contrast, the isolated *Dm* PAN3 PK and C-term domains were monomeric in solution (Figure 1F and Table S1). Thus, the PK+C fragment mediates PAN3 homodimerization, and this function is conserved from yeast to humans.

PAN3 Proteins Have Retained the Ability to Bind ATP

Given the importance of the PK and C-term domains in the assembly of the PAN2-PAN3 complex, we sought to determine the structure of a PAN3 fragment encompassing these domains. We initially expressed the corresponding fragments of the *Nc*, *Dm*, and *Hs* PAN3 proteins in *E. coli*. As protein kinases typically bind adenosine nucleotides (Taylor and Kornev, 2011), we examined the ATP-binding ability of the PAN3 proteins. Thermal-shift assays indicated that the *Nc*, *Dm*, and *Hs* PAN3 protein fragments were stabilized by the addition of ATP in a Mg^{2+}-dependent manner, while stabilization by AMP and ADP was independent of Mg^{2+} (Figure S1A and Table S2). By contrast, PAN3 proteins were not stabilized by GTP (Table S2). Consequently, protein crystallization experiments routinely included nonhydrolyzable ATP analogs and $MgCl_2$ in the protein solution.

Structure Determination of the *Nc* and *Dm* PK and C-Terminal Domains

Crystals of *Nc* PAN3 (residues 234–656, wild-type or mutant) and *Dm* PAN3 (residues 349–790) were obtained in the presence of nonhydrolyzable ATP analogs (ATP-γ-S and AMP-PNP for the *Nc* and *Dm* proteins, respectively) and $MgCl_2$. However, no crystals were obtained for the human protein in diverse conditions. Experimental phases for *Nc* PAN3 selenomethionine-substituted crystals were obtained using the single-wavelength anomalous dispersion (SAD) method. The structure of *Nc* PAN3 was refined at 3.3 Å resolution (Table 1) and was then used as a molecular replacement model to phase the structure of *Dm* PAN3 and mutant *Nc* PAN3 (*Nc* PAN3^{M}), which were refined at 3.6 Å and 2.85 Å resolution, respectively (Table 1). The *Nc* PAN3^{M} protein and other *Dm* and *Hs* PAN3 mutants were generated after the refinement of the wild-type proteins in order to promote alternative crystal packing arrangements (see Table S3 for a description of *Nc* PAN3^{M}). Higher-resolution data were obtained only for the *Nc* PAN3^{M} protein, although it crystallized in the same crystal form as wild-type *Nc* PAN3 (Table 1). Importantly, *Nc* PAN3^{M} retains the ability

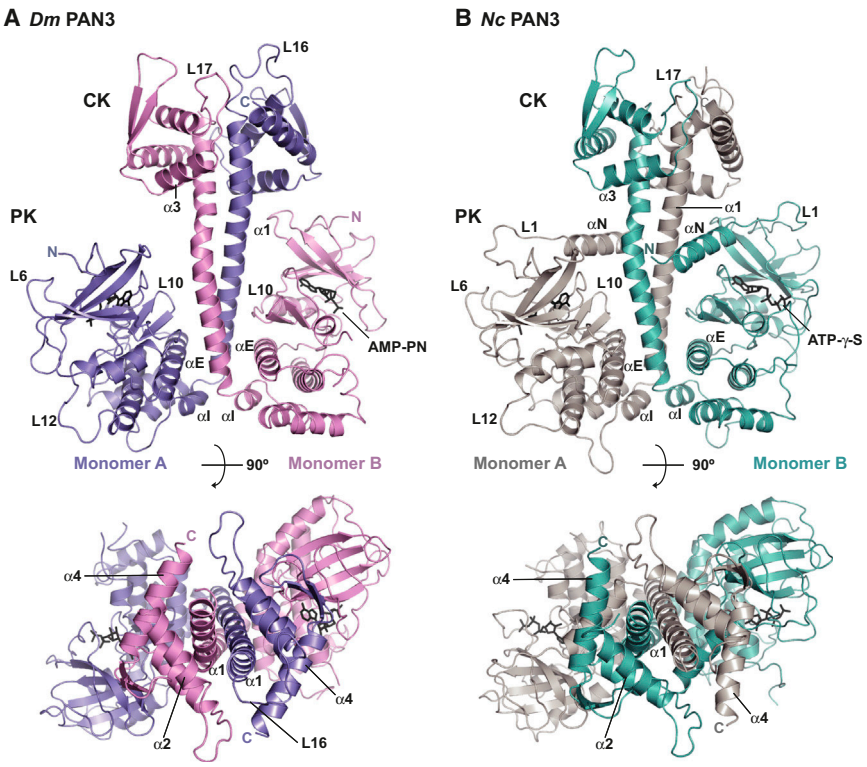


Figure 2. The Structure of PAN3 Homodimers

(A) Cartoon representation of the overall structure of the *Dm* PAN3 dimer. Chains A and B are shown in blue and magenta, respectively. The positions of the nucleotide binding sites are marked by AMP-PN shown in stick representation. Secondary structure elements are indicated. The two views are related by a rotation of 90° around the horizontal axis.

(B) Structure of *Nc* PAN3^{M} in the same orientations as in (A). Chains A and B are shown in gray and cyan, respectively. The positions of the nucleotide binding sites are marked by ATP-γ-S shown in stick representation. See also Figures S1–S3 and Table S2.

to bind ATP and has similar melting temperatures compared to the wild-type protein (Table S4).

The *Nc* and *Dm* PAN3 crystals contain one homodimer per asymmetric unit (Figure 2). Most residues were visible in the electron density, with the exception of two surface loops in both the *Nc* and *Dm* structures (*Nc* PAN3 residues 368–377 and 558–567, *Dm* PAN3 residues 488–513 in both monomers, and 700–710 in monomer B) and an N-terminal extension in the *Dm* PK domain (residues 349–378; αN and L1, see below, Figure 3A). Furthermore, clear density for ATP-γ-S was visible in the *Nc* PAN3 structures (Figure 3D). In the *Dm* PAN3 crystals, clear density for AMP-PN was observed (Figure S2A), suggesting that the γ-phosphate of the AMP-PNP analog was hydrolyzed, most likely due to the length of time (1 week) required for the *Dm* PAN3 crystals to grow.

PAN3 Forms Intertwined Dimers

PAN3 forms dimers in solution and in the crystal environment (Figures 1F and 2). Each PAN3 monomer consists of three structural domains: the PK domain at the base of the assembly, a C-terminal knob (CK) domain, and a long central α-helix (α1) that connects these two globular domains (Figures 2A and 2B). The interactions responsible for dimerization are predominantly found within the long central α1 helix, which interacts with the corresponding central α1 helix of the second monomer to form an asymmetric left-handed parallel coiled coil (Figure 2). The asymmetry is caused by a kink in only one chain of the PAN3 dimer (Figure 2 and Figure S3). As a result of the dimerization via the intertwined coiled coil, the CK domain of monomer A is positioned close to the PK domain of monomer B (Figures 2A and 2B).

Comparison of the wild-type *Nc* and *Dm* PAN3 structures indicates that the homodimers superpose well with RMSD values of 1.87 Å for monomer A and 2.94 Å for monomer B (Figure 2 and Figure S2D). The kink angle of the α1 helix is smaller in the *Nc* PAN3 protein (ca. 17°) compared to *Dm* PAN3 (ca. 22°), explaining the higher rmsd values obtained for the superposition of the B monomers (Figure S3). Comparison of the wild-type and mutant *Nc* PAN3 structures reveals high structural similarity (Figure S2E, rmsd values of 0.85 Å and 1.0 Å for monomers A and B, respectively).

One clear difference between the *Dm* and *Nc* PAN3 structures is the presence of an additional N-terminal helix preceding the *Nc* PK domain (helix αN, Figure 2A versus Figure 2B), which is arranged almost perpendicular to the central coiled-coil helices. Secondary structure prediction identified an analogous helix in the *Dm* PAN3 and *Hs* PAN3 sequences, although this region is less conserved among orthologs (Figure 3A). The similarity between the *Nc* and *Dm* PAN3 structures, together with the conservation of critical structural residues, indicates that the fold observed in these structures is likely to be common to all PAN3 family proteins.

The PAN3 Pseudokinase Domain Adopts a Divergent Kinase Fold

The PK domain of PAN3 exhibits the canonical bilobal protein kinase fold, with a smaller N-terminal lobe (N-lobe) and a larger C-terminal lobe (C-lobe) (Figures 3A and 3B). The N-lobe (*Dm* residues 383–479 and *Nc* residues 263–359) comprises a five-stranded antiparallel β sheet (β1–β5) and a flanking α helix (helix αC) (Figures 3A and 3B). The C-lobe (*Dm* residues 480–655 and *Nc* residues 360–514) is composed predominantly of α helices (Figures 3A and 3B, denoted as αD–αI). Despite retaining the general structural characteristics of protein kinases, analysis of PAN3 amino acid sequences demonstrates that the PK domain has substitutions in all the conserved motifs that are critical for kinase activity, such as in the catalytic VAIK and HRD motifs and in the Mg^{2+} binding DFG motif (Figure 3A).

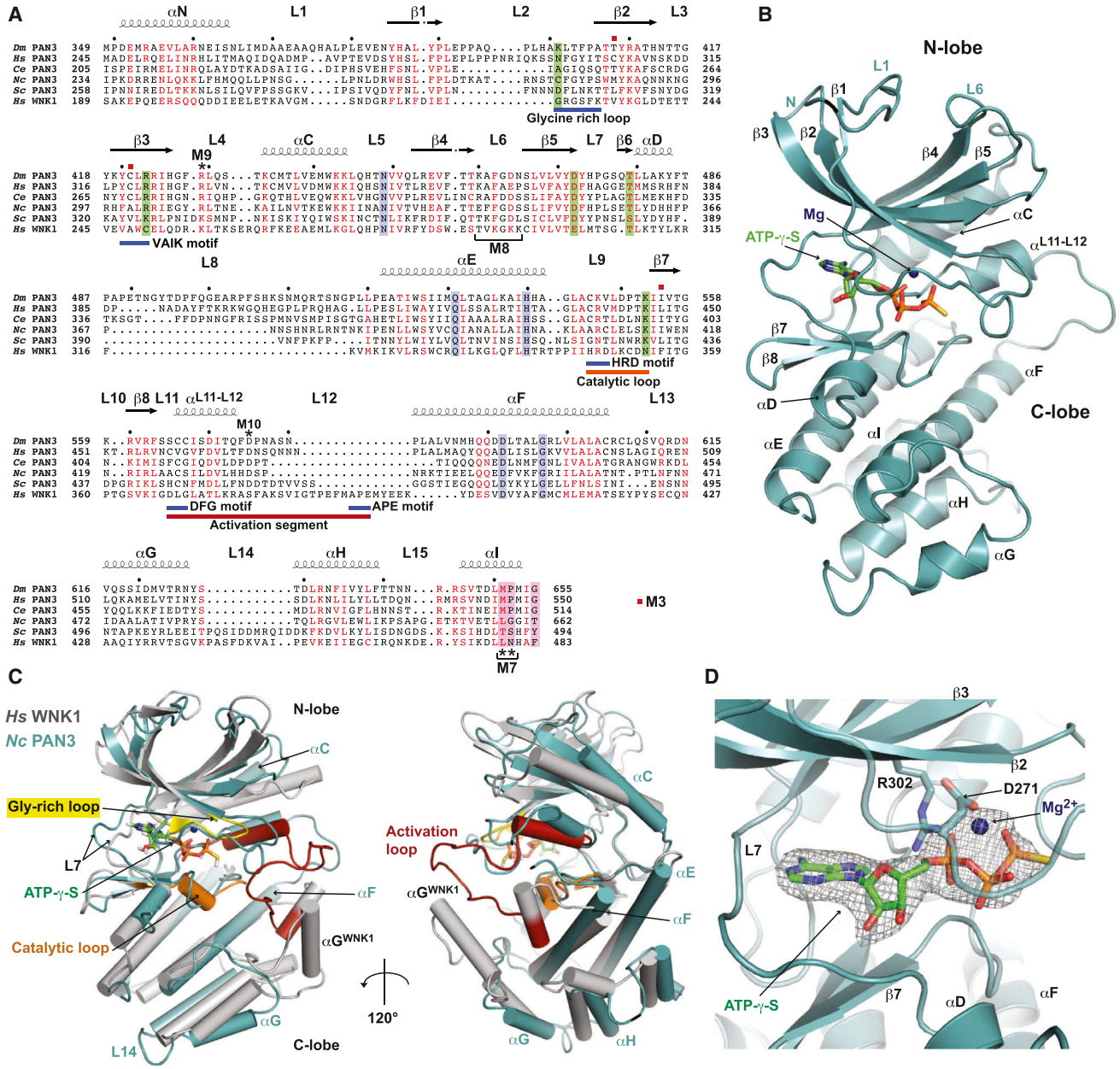


Figure 3. The PAN3 Pseudokinase Domain Binds ATP

(A) Structure-based sequence alignment of PAN3 orthologs and human WNK1. Secondary structure elements as determined from the Nc and *Dm* PAN3 structures are shown above the alignment. Conserved kinase motifs are shown below the WNK1 sequence and are as described for the canonical PKA kinase (Taylor and Kornev, 2011). Residues conserved in all aligned sequences are shown with a purple background, and residues with >70% similarity are in red. *Nc* PAN3^{M} residues that form hydrogen bonds with the ATP-γ-S molecule are shown in green background, and residues that line the W-binding pocket have a magenta background. Residues mutated in the M7, M8, M9, and M10 mutants are marked by asterisks. Residues that line the nucleotide-binding pocket that are substituted in the M3 mutant are marked by red squares. The species abbreviations are as follows: *Dm* (*Drosophila melanogaster*), *Hs* (*Homo sapiens*), *Ce* (*Caenorhabditis elegans*), *Nc* (*Neurospora crassa*), and *Sc* (*Sacchromyces cerevisiae*).

(B) Cartoon representation of *Nc* PAN3^{M} PK domain with bound ATP-γ-S and Mg^{2+} in the nucleotide-binding pocket (shown in stick and sphere representation, respectively).

(C) Superimposition of WNK1 kinase domain (gray) onto the PK domain of *Nc* PAN3^{M} (cyan). WNK1 catalytic, activation, and glycine-rich loops are highlighted in orange, red, and yellow, respectively.

(D) Close-up view of the *Nc* PAN3^{M} nucleotide binding site, with the simulated annealing $F_o - F_c$ omit map contoured at 3σ shown in gray mesh. The ATP-γ-S and Mg^{2+} molecules are overlaid, shown in stick and sphere representation, respectively. See also Figures S1 and S2 and Tables S1, S2, and S4.

Structural superposition of *Nc* and *Dm* PAN3 PK domains with all structures deposited into the Protein Data Bank (PDB) using the DALI server indicates that the WNK1 (with no Lys [K] 1) kinase is most similar in structure to the PAN3 PK domain (lowest Z score of 21.6 and 21.2 for *Dm* PAN3 and *Nc* PAN3, respectively; Figures 3A and 3C). As observed in the structure of WNK1 kinase (Min et al., 2004), the structure of PAN3 N-lobe exhibits an almost closed β-barrel appearance due to extensions at the N terminus (L1) and the loop L6 (Figures 3A and 3B).

The most striking structural difference between the PAN3 PK domain and canonical kinases is the lack of the activation loop, which normally bridges the DFG motif and the APE motif (Figures 3A and 3C). The conformation of this loop is typically regulated by phosphorylation and provides a platform for peptide substrates to bind and access the catalytic site of the kinase (Huse and Kuriyan, 2002; Taylor and Kornev, 2011). Sequence comparison shows that the activation loop and the APE motif are missing in all PAN3 orthologs (Figure 3A).

Furthermore, the substrate-binding groove, typically located between the αD and αG helices of canonical kinases, is disrupted in PAN3. This is due to a shift of the αG helix to a position closer to the N terminus of the αH helix (Figures 3B and 3C), a conformational change that likely results from a deletion in the L14 loop (Figure 3A), as observed in most PAN3 protein sequences.

The positioning of the αC helix is critical for the catalytic activity of protein kinases, and in the active state it adopts a closed conformation (Huse and Kuriyan, 2002; Taylor and Kornev, 2011). In PAN3, the αC helix is displaced away from the N-lobe of the protein. This conformation is reminiscent of the open conformation of kinase domain structures in the inactive state (Huse and Kuriyan, 2002; Taylor and Kornev, 2011). Lastly, the central αF helix of PAN3 is notably longer than that of typical kinase domains (Figures 3A–3C). Given these differences from the structure of canonical kinases, it is highly unlikely that PAN3 is an active kinase or participates in protein phosphorylation in a canonical manner.

The PAN3 Nucleotide-Binding Site

It is remarkable that the PAN3 PK domain binds ATP despite adopting a divergent kinase fold and lacking several key catalytic residues (Figure 3A and Table S2). Similar to other kinases, the ATP-binding site is located in the cleft between the N- and C-lobes of the kinase fold (Figure 3B). However, the ATP-binding pocket is wider than that of typical kinases, allowing the nucleotide to bind in a position that would clash with the L7 loop in canonical kinases (Figure 3D). Due to the higher resolution of the *Nc* PAN3^{M} structure, which has no substitutions within the nucleotide-binding site (Table S3), we focus our discussion of the PAN3 ATP-binding mode for the *Nc* protein.

In catalytically active kinases, the invariant Lys residue in the VAIK motif coordinates the α- and β-phosphate groups of the ATP molecule to position the nucleotide correctly within the active site for catalysis (Huse and Kuriyan, 2002; Taylor and Kornev, 2011). This interaction is further stabilized by an ionic interaction between the Lys residue and a highly conserved Glu in the αC helix of protein kinases in the active conformation (Huse and Kuriyan, 2002; Taylor and Kornev, 2011). In PAN3, the conserved Glu and Lys residues have been substituted with a hydrophobic residue and Arg, respectively (Figure 3A). Nonetheless, R302 in *Nc* PAN3 interacts with the α-phosphate of ATP-γ-S in a manner similar to the analogous Lys100Arg substitution in the VAIK motif of the STRADα PK (Figure 3D; Zeqiraj et al., 2009). The ATP binding scaffold is further degraded by substitutions in the Gly-rich loop (Figures 3A and 3C), which is typically responsible for interacting with the phosphate moieties of adenosine nucleotides.

Simulated annealing omit maps for *Nc* PAN3^{M} with the nucleotide and Mg^{2+} coordinates removed reveal clear difference density within the ATP-binding pocket (Figure 3D) that can be well described by a Mg^{2+}-coordinated ATP-γ-S molecule. In this model, an Asp residue (D271) adjacent to the Gly rich loop is positioned close to the Mg^{2+} ion, potentially replacing the substituted Mg^{2+} binding residues of the DFG motif and catalytic loop (Figure 3D and Figures S2B and S2C). Thus, the PAN3 PK domains appear to have adopted a noncanonical mechanism for binding adenosine nucleotides.

To test whether it is possible to generate a soluble *Dm* PAN3 dimer that does not bind ATP for functional studies, we substituted two hydrophobic residues that line the ATP-binding pocket (C421 and I555 in *Dm* PAN3, Figure S2A) with Phe residues. These mutations did not affect dimerization (Table S1). Thermal shift assays performed in the presence and absence of nucleotide and Mg^{2+} indicated that the apparent melting temperature of the double C421F,I555F mutant did not change (Figure S1B and Table S4), suggesting that the mutations abolish nucleotide binding.

We next tested whether *Dm* PAN3 has also retained the ability to hydrolyze ATP in vitro. No intrinsic ATPase activity was observed for the *Dm* PK+C fragment or the I555F and C421F,I555F mutants (which were used as negative controls), even after prolonged incubation (20 hr, Figure S4A). Under the same conditions, calf intestinal alkaline phosphatase exhibited robust ATPase activity, although its concentration was 3,000-fold lower.

The PAN3 Central Coiled Coil Mediates Dimerization

Although it is known that PAN3 oligomerizes via the PK+C-term domains (Figure 1 and Mangus et al., 2004), based on sequence alone it was difficult to predict that the self-interaction would be mediated via a coiled coil. Coil-forming helices are characterized by a seven-residue (heptad) repeat. The residues within a heptad repeat are typically denoted (a-b-c-d-e-f-g) with apolar "a" and "d" core residues located at the helical interface. The asymmetry of the PAN3 coiled coil is most likely due to the presence of highly conserved Glu and Gly residues at the "d" and "a" positions of the third and fourth heptad repeats, respectively, which disrupt the periodicity of the coiled coil (Figures 4A and 4B).

The interface between the monomers is extensive and buries a total surface area of ~2,800 and ~3,300 Å^2 for *Dm* and *Nc* PAN3, respectively. The interactions extend beyond the hydrophobic coiled-coil interface and include salt bridges, hydrogen bonds, and hydrophobic interactions between residues of the central α helix and the PK domain (Figure S2F). Additional hydrophobic interactions are also observed between the central α helix and the CK domain of the protein.

To test whether PAN3 dimerizes in vivo in a manner similar to that observed in the crystal structure, we introduced acidic

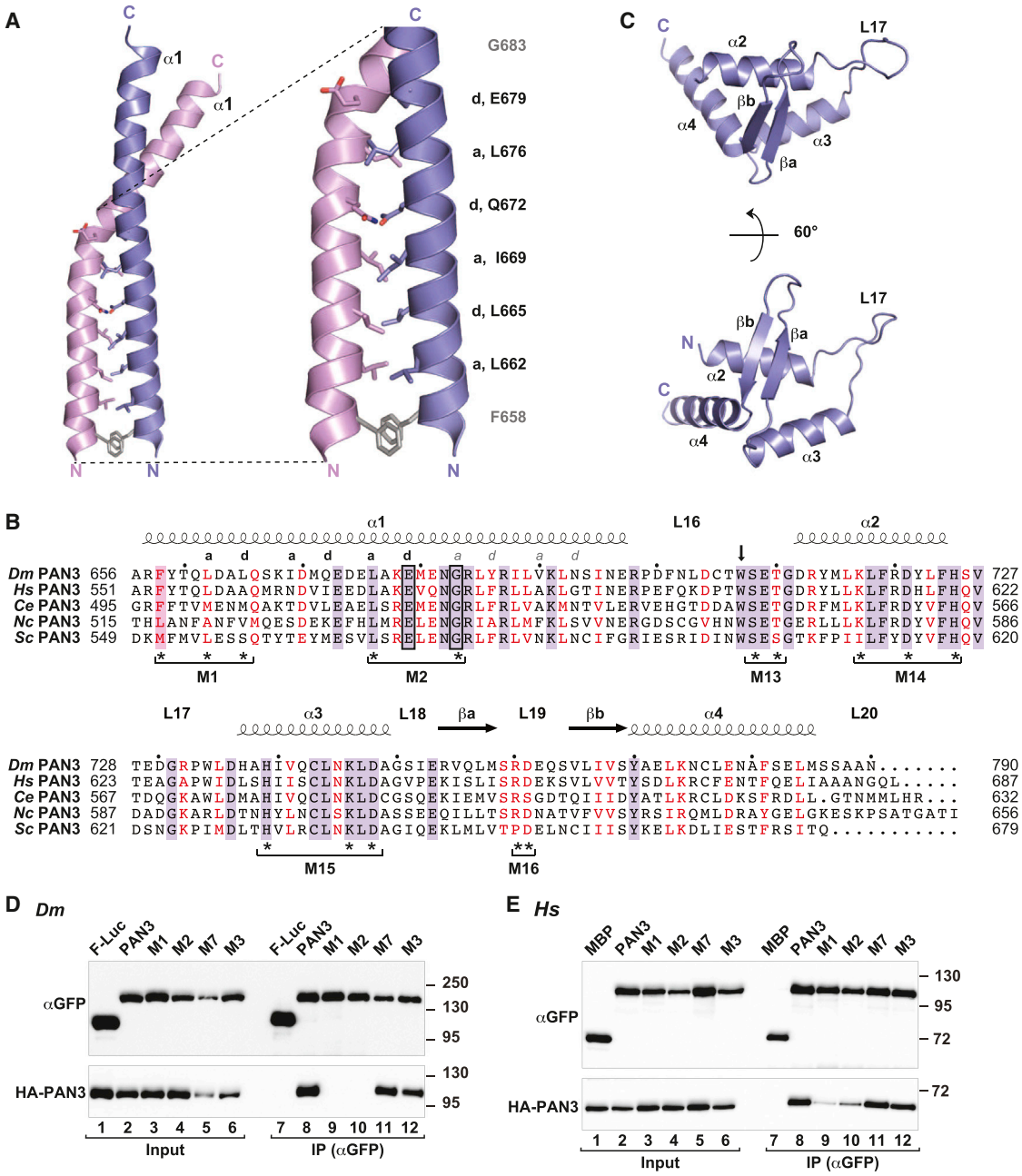


Figure 4. The Coiled-Coil Domain Mediates Homodimerization

(A) Cartoon representations of the coiled-coil domain of *Dm* PAN3. Residues corresponding to the "a" and "d" positions of the coiled-coil heptad repeat are indicated.

(B) Sequence alignment of the PAN3 C-term domain. Coiled-coil heptad residues identified in the *Dm* PAN3 structure corresponding to the "a" and "d" positions are indicated. The "a" and "d" heptad positions that cause the helical kink are boxed. Potential "a" and "d" heptad positions C-terminal to the kink are shown in gray italics. Residue W707 is indicated by an arrow. Symbols are as described in Figure 3A.

(C) Cartoon representation of the *Dm* PAN3 CK domain shown in two orientations.

(D and E) The interaction of GFP-PAN3 (wild-type or mutants) with HA-PAN3 was analyzed by coimmunoprecipitation in *Dm* S2 and human HEK293 cells as described in Figure 1, except that GFP-*Hs* PAN3 was used. GFP-MPB served as negative control in human cells. See also Figures S2 and S4 and Tables S1, S3, and S4.

residues at the "a" and "d" positions of the coiled coil and performed coimmunoprecipitation assays. The substitutions were introduced in the context of full-length *Dm* and *Hs* PAN3 proteins. Specifically, we generated a triple mutant (M1) and a double mutant (M2) (Figure 4B and Table S3). These mutations were sufficient to disrupt *Dm* PAN3 homodimerization in vitro

and in cells (Figure 4D, lanes 9 and 10) without affecting the protein fold as indicated by MALLS analysis and thermal shift assays (Tables S1 and S4). The analogous mutations also impaired the ability of *Hs* PAN3 to self-interact in vivo (Figure 4E, lanes 9 and 10). Although the *Hs* PAN3 M1 mutant was monomeric in solution (Table S1), MALLS analysis indicated that the M2 mutant was in part monomeric and in part dimeric, suggesting that this mutation is not sufficient to completely abolish dimerization of the human protein (Table S1). Importantly, the integrity of the ATP-binding pocket was not required for *Dm* or *Hs* PAN3 dimer formation (Mutant M3, Figures 4D and 4E, lanes 12; and Tables S1 and S3).

Structure of the C-Terminal Knob Domain

Despite sequence conservation, the architecture of the PAN2-binding domain of PAN3 has remained enigmatic due to the lack of sequence homology with known structural folds determined to date. The structures presented in this study now reveal that the PAN2 binding domain of PAN3 can be divided into the coiled-coil and the CK domains. The CK domain adopts a compact conformation comprised of three α helices (α2, α3, and α4) connected by loop L17 and a β-hairpin inserted between α helices α3 and α4 (Figures 4B and 4C). We compared the *Dm* CK domain with the structures deposited in the PDB using the DALI server. Some helix-turn-helix (HTH) domains have remote similarity to the CK domain (Figure S4B). However, the highest Z score was 3.3, suggesting that the PAN3 CK domain adopts a rather unique protein fold.

Identification of the PAN2-Binding Interface

PAN2 interacts with the PAN3 C-term fragment, which remains monomeric in solution (Figure 1F), indicating that PAN2 binding does not require PAN3 dimerization. Accordingly, we observed that none of the aforementioned dimerization mutants (M1 and M2, Table S3) disrupted PAN3 interaction with PAN2 in either *Dm* or human cells (Figures S4C and S4D).

Although PAN2 binding to PAN3 does not require the PK domain, we wondered if nucleotide binding could modulate this interaction. We observed that the mutations of the ATP-binding pocket alone (mutant M3) or combined with the M2 dimerization mutations (mutant M2+M3) did not affect PAN3-PAN2 interaction in *Dm* S2 cells (Figure S4C and Table S3). Together these results confirm that dimerization and nucleotide binding are not required for PAN3 to interact with PAN2.

Next, we analyzed the surface conservation of the C-term domain to identify surface-exposed residues that could mediate the PAN2-PAN3 interaction (Figure 5A). This analysis identified a number of highly conserved exposed residues clustered on the surface of the *Dm* PAN3 CK domain (Figure 5A). The potential role of these residues in PAN2 binding was tested by mutagenesis (Mutants M13–M16; Figures 4B, 5A, and 5B and Table S3).

The residues substituted in the M13, M14, and M16 mutants are clustered on the surface of the CK domain and represent the most conserved region of the whole PAN3 dimer structure (Figures 5A and 5B). These mutations were generated in the full-length protein and were combined with the M2 (*Dm*) or M1 (*Hs*) mutations to prevent dimerization with the endogenous wild-type protein, which could confound the interpretation of the results. We observed that when combined with M1, the M13, M15, and M16 mutations strongly reduced or abolished PAN2 binding in human cells (Figure 5C). By contrast, PAN2 binding was reduced only when mutations M15+M16 and M14+M15+M16 were combined with M2 in *Dm* cells (mutants M17 and M18, respectively; Figure 5D and Table S4), whereas the isolated M13, M14, M15, and M16 mutations (in combination with M2) were ineffectual in inhibiting binding (Figure S4E). Collectively, these results identify the binding surface of PAN2 on the PAN3 CK domain.

The PAN3 Dimerization Interface Harbors a W-Binding Pocket

Unexpectedly, crystal-packing interactions within the *Dm* PAN3 lattice revealed the presence of a W-binding pocket at the base of the coiled coil (Figures 6A and 6B). This hydrophobic pocket lies in a special position on the pseudo-two-fold axis of the coiled coil that relates two PAN3 monomers (Figures 6A–6C). Therefore, the pocket is lined by residues M651, P652, G655, and F658 from monomer A and their symmetry mates from monomer B (Figures 3A and 4B). The pocket accommodates the exposed W707 (loop L16) from monomer A of a neighboring molecule in the crystal, such that the indole ring deeply stacks between two symmetry-related proline residues (P652) located at the base of the coiled-coil domain (Figures 6B).

Notably, W707 is flanked by T706 and S708. T/SW motifs are frequently found in the silencing domains (SDs) of TNRC6 proteins and have previously been shown to contribute to PAN3 binding (Chekulaeva et al., 2011; Huntzinger et al., 2013). We therefore reasoned that the observed interaction might mimic a bona fide interaction between PAN3 and TNRC6 proteins. Consistent with this hypothesis, the pocket-forming residues are conserved in higher eukaryotes where TNRC6 orthologs have been identified (Figure 3A). By contrast, the pocket is absent in the *Nc* PAN3 structure and appears to be absent in most fungi, which lack GW182 orthologs.

Importantly, the polypeptide chain around W707 is ordered only in the *Dm* PAN3 monomer that interacts with the pocket, demonstrating a disorder-to-order transition upon binding similar to that observed for disordered short-linear motifs (Tompa, 2012). A similar mode of binding is expected for TNRC6-derived T/SW peptides binding to the PAN3 surface. Notably, the backbone between T706 and D712 is clearly defined in the *Dm* PAN3 density, with the bulky side chain of W707 acting as a sequence anchor and defining the register.

Although the low resolution of the *Dm* PAN3 structure does not allow the detailed analysis of hydrogen bonds, it is clear that the ordered flanks of W707 form additional contacts with the surface of the PAN3 dimers (Figure 6C). We therefore assume that the flanking T/S residues with their general capacity to stabilize induced loop conformations and to form intermolecular contacts are of additional importance in this context. Notably, we observed that alanine substitution of Trp residues in the M2 region of the TNRC6C-SD, which are mostly flanked by T/S residues, abolished PAN3 binding in human cells (Figure 6D and Figure S5A). By contrast, substitutions in the CCR4-interacting motifs 1 and 2 (CIM-1 and CIM-2) did not prevent PAN3 binding (Figure 6D, Figure S5A, and Table S3). Thus, the W-containing

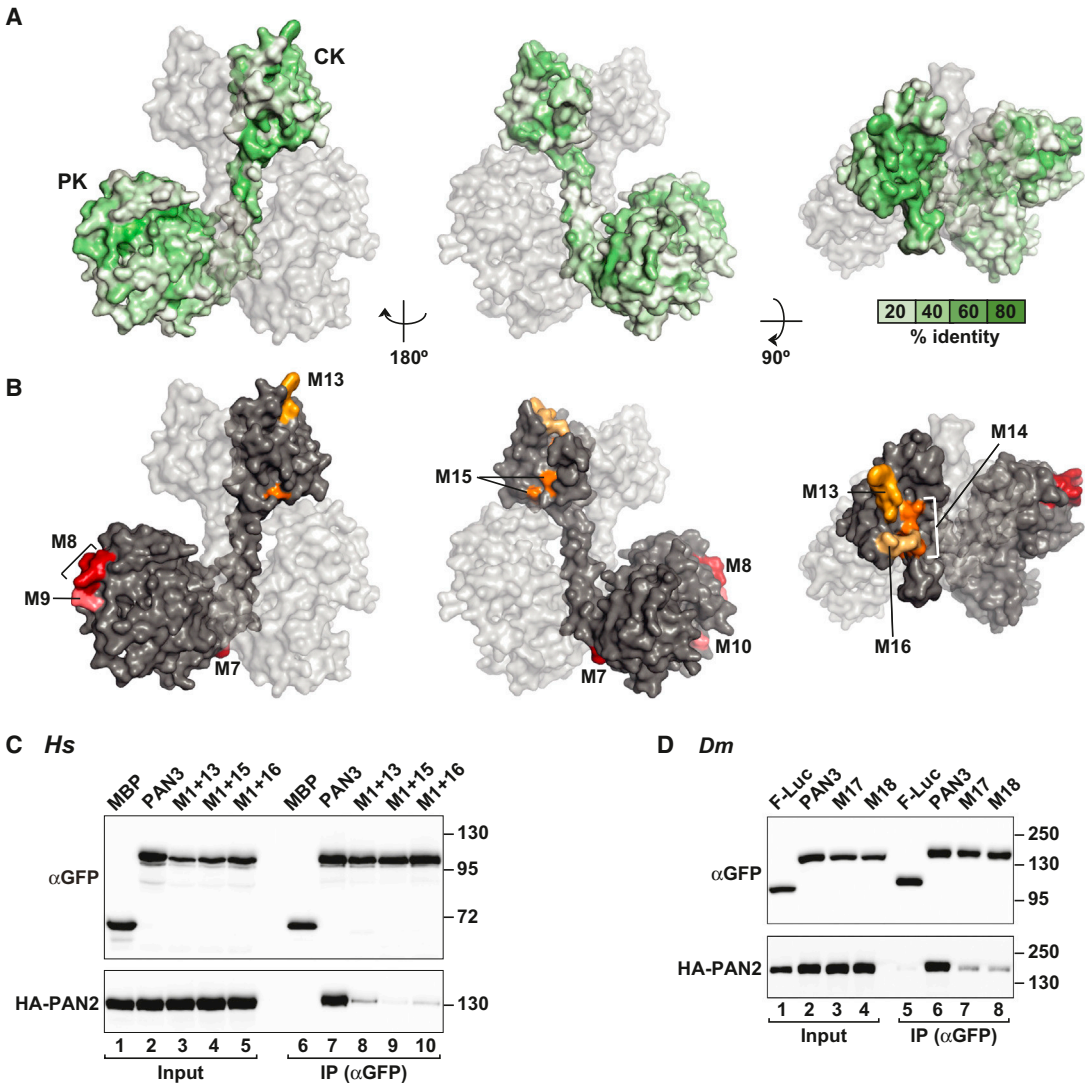


Figure 5. The PAN2 Binding Surface

(A) Surface representation of *Dm* PAN3 colored by sequence conservation on monomer A (monomer B is shown in transparent gray surface representation).

(B) Surface representation of *Dm* PAN3, shown in the same orientations as in (A), with mutated residues introduced in this study highlighted. Only mutated residues of monomer A are shown. Mutations in the PK and C-term domains are shown in red and orange, respectively.

(C and D) The interaction of GFP-PAN3 (wild-type or mutants) with HA-PAN2 was analyzed in *Dm* and human cells as described in Figure 1. See also Figure S4 and Table S3.

motifs in the M2 region of TNRC6C provide a major contribution to PAN3 binding.

To investigate if the W-binding pocket is indeed relevant for TNRC6 binding, we substituted *Hs* PAN3 M546 (*Dm* PAN3 M651) with Phe to fill the cavity with a bulky side chain and P547 (*Dm* PAN3 P652) with Gly to prevent stacking interactions (mutant M7, Figures 3A and 5B, and Table S3). Remarkably, these mutations strongly reduced *Hs* PAN3 interaction with the TNRC6C-SD in vivo and in vitro (6C-SD, Figures 6E and 6F). Single M546F and M546H substitutions also reduced binding of the TNRC6C-SD in vitro in GST pull-down assays (Figure S5B). PAN3 dimerization and binding to PAN2 were not affected by the M7 mutation (Figures 4D and 4E, lanes 11; and Figures S4C and S4D), indicating that the M7 mutant is properly folded. Furthermore, MALLS analysis and thermal shift assays indicate that the *Hs* PAN3 M7 mutant behaves as a well-folded dimer that is capable of binding ATP in vitro (Tables S1 and S4).

Because the W-binding pocket forms at the dimer interface, we expected that the PAN3 dimerization mutants would also affect TNRC6C binding, in a manner similar to that of the M7 mutant. Consistent with this expectation, the dimerization mutants (M1 and M2) were strongly impaired in TNRC6C-SD binding (Figure 6E). Importantly, the *Hs* PAN3 M7 mutation also reduced the interaction with the silencing domain of TNRC6B (Figure 6G). Similar results were obtained when single or double mutants of the *Dm* PAN3 protein were tested for binding to

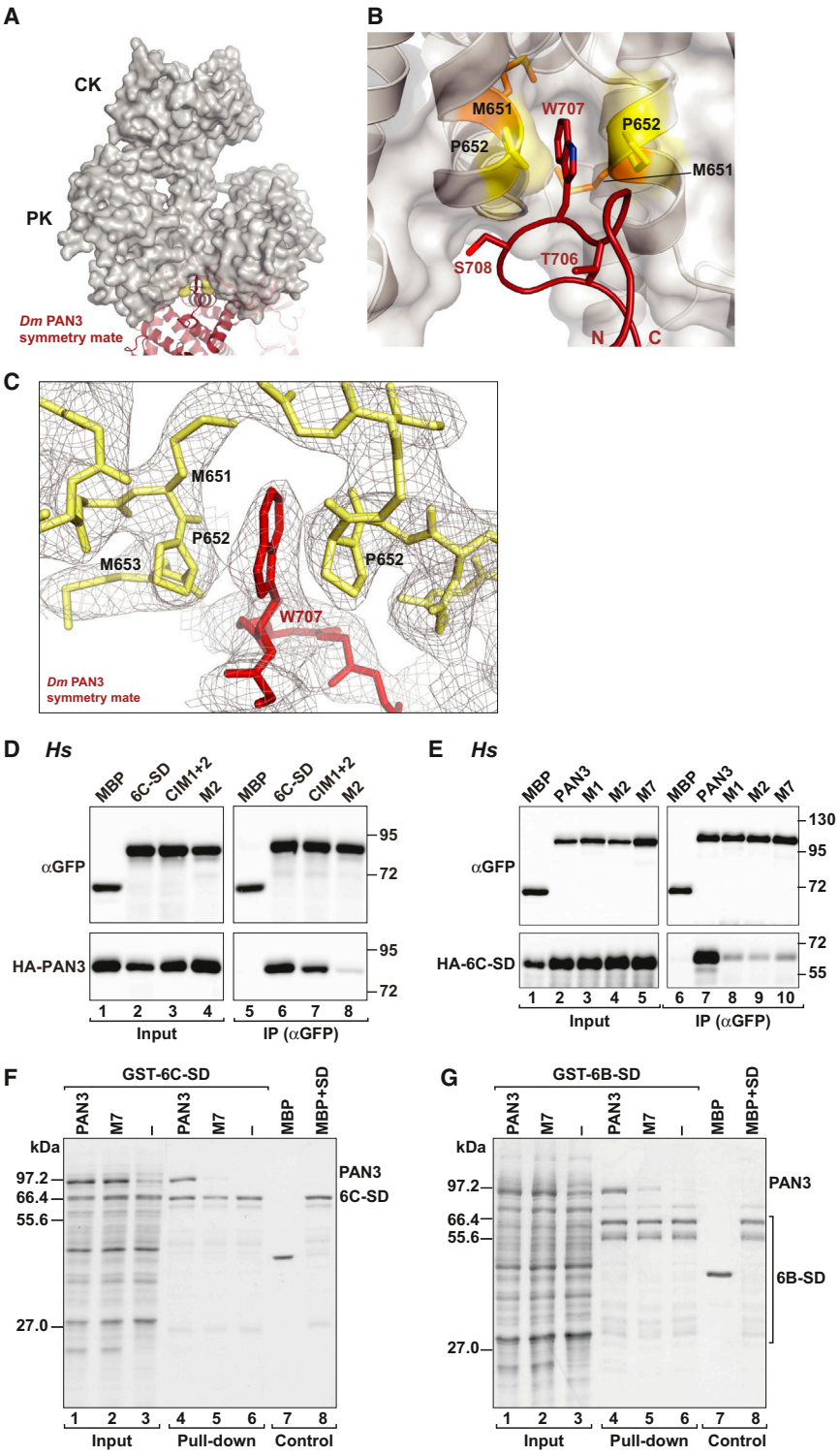


Figure 6. A W-Binding Pocket at the Dimeric *Dm* PAN3 Interface Is Required for Binding to TNRC6 Proteins

(A) Surface representation of *Dm* PAN3 homodimer. The binding pocket at the base of the molecule formed by the symmetric proline residues is highlighted in yellow, and monomer A of a symmetry-related *Dm* PAN3 homodimer is shown in red.

(B) Close-up view of the W-binding pocket. The *Dm* PAN3 homodimer is shown as a transparent surface with cartoon representation underneath. A loop harboring T706-W707-S708 from monomer A of a symmetry-related *Dm* PAN3 homodimer is shown as cartoon and stick representation (colored in red).

(C) Final $2F_o - F_c$ density for the W-binding pocket and for the L16 loop is shown as a gray mesh and contoured at 1σ. The *Dm* PAN3 homodimer is shown as yellow sticks, and the L16 loop of a symmetry mate is colored in red. For clarity, some side chains and residues have been omitted.

(D) Interaction between GFP-TNRC6C-SD (6C-SD, wild-type, or mutant) and HA-*Hs* PAN3.

(E) Interaction of GFP-*Hs* PAN3 (wild-type or mutant) with the silencing domain of TNRC6C.

(F and G) GST pull-downs using recombinant MBP-*Hs* PAN3 (wild-type or M7 mutant) and GST-6C-SD (F) or GST-6B-SD (G). MBP served as a negative control. See also Figure S5 and Table S3.

TNRC6B-SD in vitro (Figure S5C). This interaction was tested because TNRC6 proteins complement silencing in *Dm* cells (Huntzinger et al., 2013). We conclude that the W-binding pocket represents a major binding site for TNRC6B and TNRC6C on the PAN3 protein.

PAN2 Binding and the Integrity of the Nucleotide Binding Pocket Are Required for mRNA Degradation

To evaluate the relevance of PAN3 dimerization, nucleotide binding, and PAN2 interaction in mRNA degradation, we compared the activity of *Dm* PAN3 mutants with the wild-type protein using a well-established mRNA degradation assay in *Dm* S2 cells. This assay involves the expression of a λN-HA-fusion of PAN3 that binds with high affinity to five consecutive BoxB sites (5BoxB) in the 3′UTR of a firefly luciferase (F-Luc) reporter mRNA (Braun et al., 2011).

Tethered *Dm* PAN3 reduced steady-state mRNA levels; this activity was only slightly affected by mutations that disrupt dimerization (M1, M2, Figures 7A and 7B). To determine whether the reduction on mRNA levels was due to mRNA destabilization, we measured mRNA half-lives. In control cells, the half-life of *F-Luc-5BoxB* mRNA was 153 ± 8 min (Figure 7C). Expression of wild-type PAN3 or the M2 dimerization mutant accelerated the degradation of the mRNA reporter, resulting in shorter mRNA half-lives (20 ± 4 min and 32 ± 7 min, respectively; Figures 7D and 7E).

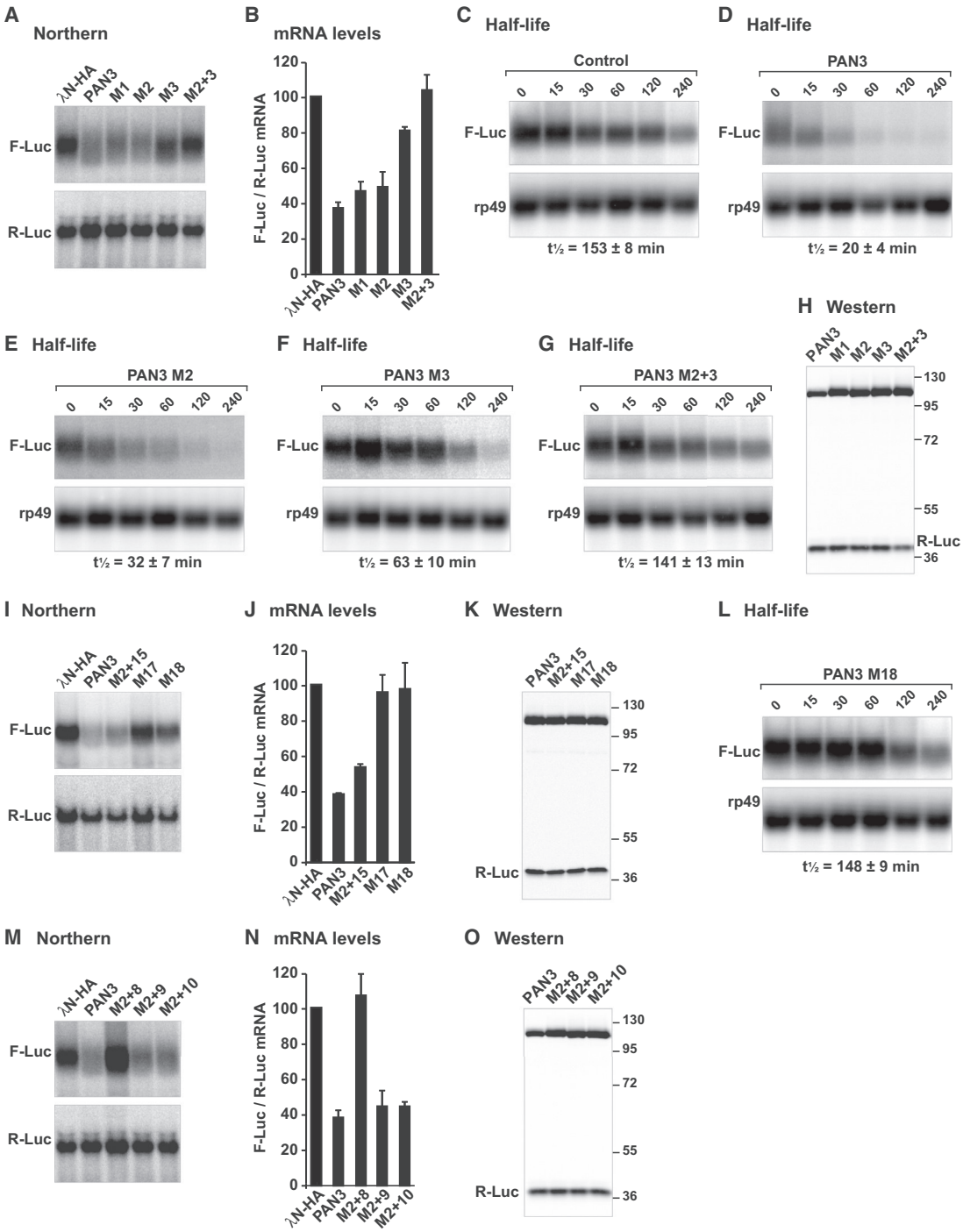


Figure 7. The Integrity of the Nucleotide Binding Site and PAN2 Binding Are Required for mRNA Degradation

The decay of the *F-Luc-5BoxB* mRNA reporter was monitored at steady state and following transcription inhibition by actinomycin D in control cells (expressing the λN-HA peptide) and in cells expressing PAN3 (either wild-type or the indicated mutants).

(A, I, and M) Northern blot analysis of representative RNA samples at steady state.

(B, J, and N) The levels of the *F-Luc-5BoxB* mRNA were normalized to that of the transfection control (*R-Luc* mRNA). The values were set at 100% in control cells expressing the λN-HA peptide. Bars represent mean values, and error bars represent standard deviations from at least three independent experiments.

(C, D, E, F, G, and L) The panels show northern blot analysis of representative RNA samples. The levels of *F-Luc-5BoxB* mRNA were normalized to the levels of long-lived *rp49* mRNA and were plotted against time. The mRNA half-lives ($t_{1/2}$) ± standard deviations calculated from the decay curves (data not shown) obtained from three independent experiments are indicated below the panels.

(H, K, and O) Western blots showing comparable expression of *Dm* PAN3 wild-type and mutants. See also Figure S6 and Table S3.

Remarkably, mutations that disrupt the nucleotide-binding pocket (M3) impaired *Dm* PAN3 activity in tethering assays, resulting in an mRNA half-life of 63 ± 10 min (Figures 7A, 7B, and 7F). The activity of the M3 mutant was further decreased when it was combined with mutations that prevent dimerization (M2+3; Figures 7A and 7B), resulting in an mRNA half-life (141 ± 13 min) comparable to that observed in the absence of PAN3 (Figure 7G versus Figure 7C). All proteins were expressed at comparable levels (Figure 7H). These results indicate that the integrity of the nucleotide-binding pocket is required for PAN3-mediated mRNA degradation.

Importantly, the mutations that impair PAN2 binding (M17 and M18) also abolished PAN3 activity in this assay (Figures 7I and 7J), although the mutants were expressed at levels comparable to those of wild-type PAN3 (Figure 7K). For example, in cells expressing the M18 mutant, the half-life of the *F-Luc-5BoxB* mRNA (148 ± 9 min) was similar to that observed in control cells (Figure 7L versus Figure 7C). This result further demonstrates that the degradation of the mRNA reporter is mediated by the PAN2-PAN3 complex.

The impact of mutations in the nucleotide-binding pocket on PAN3 activity was surprising because this mutation did not affect PAN2 binding (Figure S4C, lane 12). These results suggest that although the PK domain does not contribute to PAN2 binding, it may exert a regulatory function. To identify additional residues in the PK domain that could potentially regulate mRNA degradation, we analyzed the surface conservation of the domain and identified a number of conserved surface-exposed residues (Figure 5A). The potential role of these residues in mRNA degradation was tested by mutagenesis. In particular, we deleted the surface loop L6, which connects strands β4 and β5 within the N-lobe of the PK domain (mutant M8; Figures 3A and 5B and Table S3). We also introduced charge reversal mutations of highly conserved residues (mutant M9 and M10). These mutations are clustered on one side of the PK domain and were combined with M2 to abrogate the contribution of the endogenous wild-type protein.

Remarkably, deletion of the conserved surface loop L6 in the PK domain (M8) abolished *Dm* PAN3 activity in tethering assays. By contrast, the M9 and M10 mutations were ineffectual (Figures 7M and 7N). The mutant proteins were expressed at levels comparable to wild-type (Figure 7O). Because the M8 deletion does not affect PAN2 binding (Figure S4F), these results suggest that mRNA degradation by the PAN2-PAN3 complex requires conformational rearrangements or the interaction with additional partners that are not possible with the M8 deletion. Importantly, neither PAN3 nor the mutants affected the expression of a F-Luc reporter lacking the BoxB elements (Figure S6), indicating that PAN3 must bind to the mRNA reporter in order to promote its degradation.

DISCUSSION

Although PAN2 provides the catalytic activity for deadenylation, PAN3 is key in orchestrating the interactions that recruit PAN2 to mRNA targets. This is achieved through interactions with PABP and GW182/TNRC6 proteins in general and miRNA-mediated mRNA degradation, respectively (Mangus et al., 2004; Siddiqui et al., 2007; Uchida et al., 2004; Braun et al., 2011; Chekulaeva et al., 2011; Fabian et al., 2011; Huntzinger et al., 2013). Through a combined structural and functional analysis, we identified critical residues that mediate PAN3 interaction with PAN2 and TNRC6 proteins. Our study provides a detailed molecular model for the assembly of the PAN2-PAN3 complex that significantly improves our understanding of PAN2-PAN3-mediated mRNA degradation.

PAN3 Dimerizes through an Asymmetric Coiled Coil

Our data demonstrate that PAN3 homodimerizes via an asymmetric coiled coil that connects the PAN3 PK and CK domains. The strict conservation of the Glu-Gly pair at the "d" and "a" positions of the third and fourth heptad repeat, respectively, in the PAN3 coiled coil together with the presence of the asymmetric coiled-coil conformation in both *Nc* and *Dm* PAN3 proteins and in two different crystal-packing environments indicates that asymmetric dimerization is likely a conserved feature of PAN3 proteins. Although dimerization is not required for PAN2 binding, it is clearly required to generate the W-binding pocket and thus for the recruitment of PAN2-PAN3 to miRNA targets. It is possible that other protein partners similarly use this pocket to recruit the PAN2-PAN3 complex to their targets or to regulate its activity.

Molecular Insights into the PAN2-PAN3 Interaction

PAN2 recruitment to mRNA targets is dependent on its interaction with PAN3. Characterization of this interaction is therefore essential in understanding PAN2-PAN3 function. We have demonstrated that the C-term domain of PAN3 mediates PAN2 binding in *Dm* and human cells (Figure 1) and identified solvent-exposed residues on the PAN3 CK domain that are required for this interaction (Figure 5). These results, together with the observation that PAN3 homodimerization is not required for PAN2 binding, suggest that each PAN3 monomer has a PAN2 interaction surface. However, whether it is sterically possible for two PAN2 proteins to interact simultaneously with a PAN3 dimer remains to be determined.

The W-Binding Pocket Provides a Binding Site for TNRC6 Proteins

GW182 family proteins are characterized by their richness in tryptophan residues, which are usually located in a sequence environment of predicted structural disorder. These W-containing motifs are frequently flanked either by glycine (termed GW motifs) or by serines/threonines (termed S/TW motifs), and the Trp residues in these sequence contexts have been shown to mediate the interactions of GW182 proteins with a number of protein partners via cumulative avidity effects (Fabian et al., 2011; Chekulaeva et al., 2011; Huntzinger et al., 2013).

In molecular terms, it has been speculated that the Trp residues are accommodated in hydrophobic pockets of the protein partners and that several such pockets and their spatial arrangement could confer increased affinity and specificity. This hypothesis has been confirmed by the crystal structure of human Argonaute2 (AGO2), which revealed tandem W-binding pockets occupied by free tryptophan that was included in the crystallization condition (Schirle and MacRae, 2012). However, this

structure could not address the important question of the degree to which the flanking sequences might play a role because the tryptophans were not in the context of a polypeptide.

The present structure reveals that sequences flanking tryptophan residues become ordered upon binding and can contribute to the interaction. Serine and threonine residues are particularly suited for such additional interactions, either forming direct contacts to the binding partners or stabilizing the induced conformation of the peptide backbone by internal hydrogen bonds. By contrast, flanking glycines may provide additional conformational freedom for the peptide backbone to explore surfaces of the binding partners.

Increasing evidence indicates that tryptophans can contribute to the molecular recognition of intrinsically disordered peptides in different ways. One example is the PAM2 motif of TNRC6/GW182 proteins, which binds to the C-terminal domain of PABPC and contains an additional tryptophan that adds significant binding energy (Jinek et al., 2010; Kozlov et al., 2010). However, this tryptophan is not recognized specifically, although one face of its indole ring contacts the shallow surface of PABPC. Its contribution is primarily indirect, stabilizing the bound peptide internally. By contrast, PAN3 and AGO2 rely on a different mechanism to interact with W-containing sequences. This interaction, unlike that of the TNRC6/GW182 PAM2 motif, is predominantly mediated by the Trp side chain that is directly and specifically buried within deep binding pockets, potentially assisted by flanking backbone residues to variable degrees. In the case of PAN3 it is even possible that PAN3 dimers are prone to regulated conformational flexibility. This is suggested by the presence of ATP-binding domains, as well as its asymmetric structure, which is presumably switching conformations between monomers A and B at a certain rate. This opens the exciting possibility of a regulated rather than constitutive recruitment of the PAN2-PAN3 deadenylase complex by GW182/TNRC6 proteins.

The Integrity of the PAN3 Nucleotide-Binding Pocket Is Required for mRNA Degradation

The PAN3 structures determined in this study represent one of the most conformationally divergent kinase folds determined to date. Surprisingly, PAN3 proteins are capable of binding ATP in a Mg^{2+}-dependent manner utilizing noncanonical residues for the interaction (Figures 3D and Figures S2B and S2C). Furthermore, mutations that prevent ATP binding impair degradation of an mRNA reporter in vivo (Figure 7). These results reveal an unexpected role for the integrity of the nucleotide-binding pocket in mRNA degradation. An open question is whether nucleotide binding induces conformational changes that are used to regulate deadenylation or eventually couple deadenylation to the energetic state of the cell, as suggested by studies linking glucose homeostasis and mRNA half-lives (Munchel et al., 2011).

In conclusion, the structural and functional insights into PAN3-PAN2 and PAN3-TNRC6 interactions, the identification of a conserved surface in the PK domain required for mRNA degradation, and the finding that nucleotide binding is required for the activity of the PAN2-PAN3 complex in vivo provide a framework for future studies in the elucidation of the role of this complex in posttranscriptional mRNA regulation.

EXPERIMENTAL PROCEDURES

Protein Expression and Purification

Detailed protocols for protein expression, purification, MALLS, and thermal shift assays can be found in the Supplemental Experimental Procedures.

Crystallization and Structure Determination

Crystals of *Nc* and *Dm* PAN3 were grown and data were collected, processed, and refined as described in the Supplemental Information. Diffraction data were collected on a PILATUS 6M detector at the PXII beamline of the Swiss Light Source (SLS). Native data for *Dm* PAN3 and *Nc* PAN3 were collected at 1.0 Å. SAD data were collected at the SeMet peak (0.9792 Å) for the SeMet-derivatized *Nc* PAN3 crystals. The refinement statistics are summarized in Table 1.

GST Pull-Down, Coimmunoprecipitation, and Tethering Assays

GST pull-downs, coimmunoprecipitation, and tethering assays were performed as previously described (Braun et al., 2011). Detailed protocols can be found in the Supplemental Experimental Procedures. The mutations introduced in this study are described in Table S3.

ACCESSION NUMBERS

The atomic coordinates of *Nc* PAN3 wild-type and mutant and *Dm* PAN3 are deposited in the Protein Data Bank (PDB) under ID codes 4BWK, 4BWX, and 4BWP, respectively.

SUPPLEMENTAL INFORMATION

Supplemental Information includes six figures, four tables, Supplemental Experimental Procedures, and Supplemental References and can be found with this article online at http://dx.doi.org/10.1016/j.molcel.2013.07.011.

ACKNOWLEDGMENTS

We thank the staff of the Swiss Light Source (Villigen, Switzerland) for assistance during data collection. We are grateful to D.H. Scharf and A.A. Brakhage (Hans-Knoell-Institut, Jena) for providing the *Neurospora crassa* cDNA library, C. Romier for the pNYC and pNEA vectors, and S. Helms for technical assistance. This study was supported by the Max Planck Society and by the Gottfried Wilhelm Leibniz Program awarded to E.I.

Received: March 12, 2013
Revised: June 3, 2013
Accepted: July 9, 2013
Published: August 8, 2013

REFERENCES

Boeck, R., Tarun, S., Jr., Rieger, M., Deardorff, J.A., Müller-Auer, S., and Sachs, A.B. (1996). The yeast Pan2 protein is required for poly(A)-binding protein-stimulated poly(A)-nuclease activity. J. Biol. Chem. *271*, 432–438.

Braun, J.E., Huntzinger, E., Fauser, M., and Izaurralde, E. (2011). GW182 proteins directly recruit cytoplasmic deadenylase complexes to miRNA targets. Mol. Cell *44*, 120–133.

Brown, C.E., Tarun, S.Z., Jr., Boeck, R., and Sachs, A.B. (1996). PAN3 encodes a subunit of the Pab1p-dependent poly(A) nuclease in Saccharomyces cerevisiae. Mol. Cell. Biol. *16*, 5744–5753.

Chekulaeva, M., Mathys, H., Zipprich, J.T., Attig, J., Colic, M., Parker, R., and Filipowicz, W. (2011). miRNA repression involves GW182-mediated recruitment of CCR4-NOT through conserved W-containing motifs. Nat. Struct. Mol. Biol. *18*, 1218–1226.

Fabian, M.R., Cieplak, M.K., Frank, F., Morita, M., Green, J., Srikumar, T., Nagar, B., Yamamoto, T., Raught, B., Duchaine, T.F., and Sonenberg, N. (2011). miRNA-mediated deadenylation is orchestrated by GW182 through

two conserved motifs that interact with CCR4-NOT. Nat. Struct. Mol. Biol. *18*, 1211–1217.

Hammet, A., Pike, B.L., and Heierhorst, J. (2002). Posttranscriptional regulation of the RAD5 DNA repair gene by the Dun1 kinase and the Pan2-Pan3 poly(A)-nuclease complex contributes to survival of replication blocks. J. Biol. Chem. *277*, 22469–22474.

Huntzinger, E., Kuzuoglu-Öztürk, D., Braun, J.E., Eulalio, A., Wohlbold, L., and Izaurralde, E. (2013). The interactions of GW182 proteins with PABP and deadenylases are required for both translational repression and degradation of miRNA targets. Nucleic Acids Res. *41*, 978–994.

Huse, M., and Kuriyan, J. (2002). The conformational plasticity of protein kinases. Cell *109*, 275–282.

Jinek, M., Fabian, M.R., Coyle, S.M., Sonenberg, N., and Doudna, J.A. (2010). Structural insights into the human GW182-PABC interaction in microRNA-mediated deadenylation. Nat. Struct. Mol. Biol. *17*, 238–240.

Kozlov, G., Safaee, N., Rosenauer, A., and Gehring, K. (2010). Structural basis of binding of P-body-associated proteins GW182 and ataxin-2 by the MIle domain of poly(A)-binding protein. J. Biol. Chem. *285*, 13599–13606.

Lowell, J.E., Rudner, D.Z., and Sachs, A.B. (1992). 3′-UTR-dependent deadenylation by the yeast poly(A) nuclease. Genes Dev. *6*, 2088–2099.

Mangus, D.A., Evans, M.C., Agrin, N.S., Smith, M., Gongidi, P., and Jacobson, A. (2004). Positive and negative regulation of poly(A) nuclease. Mol. Cell. Biol. *24*, 5521–5533.

Min, X., Lee, B.H., Cobb, M.H., and Goldsmith, E.J. (2004). Crystal structure of the kinase domain of WNK1, a kinase that causes a hereditary form of hypertension. Structure *12*, 1303–1311.

Munchel, S.E., Shultzaberger, R.K., Takizawa, N., and Weis, K. (2011). Dynamic profiling of mRNA turnover reveals gene-specific and system-wide regulation of mRNA decay. Mol. Biol. Cell *22*, 2787–2795.

Schirle, N.T., and MacRae, I.J. (2012). The crystal structure of human Argonaute2. Science *336*, 1037–1040.

Siddiqui, N., Mangus, D.A., Chang, T.C., Palermino, J.M., Shyu, A.B., and Gehring, K. (2007). Poly(A) nuclease interacts with the C-terminal domain of polyadenylate-binding protein domain from poly(A)-binding protein. J. Biol. Chem. *282*, 25067–25075.

Taylor, S.S., and Kornev, A.P. (2011). Protein kinases: evolution of dynamic regulatory proteins. Trends Biochem. Sci. *36*, 65–77.

Tompa, P. (2012). Intrinsically disordered proteins: a 10-year recap. Trends Biochem. Sci. *37*, 509–516.

Tucker, M., Valencia-Sanchez, M.A., Staples, R.R., Chen, J., Denis, C.L., and Parker, R. (2001). The transcription factor associated Ccr4 and Caf1 proteins are components of the major cytoplasmic mRNA deadenylase in Saccharomyces cerevisiae. Cell *104*, 377–386.

Uchida, N., Hoshino, S., and Katada, T. (2004). Identification of a human cytoplasmic poly(A) nuclease complex stimulated by poly(A)-binding protein. J. Biol. Chem. *279*, 1383–1391.

Wahle, E., and Winkler, G.S. (2013). RNA decay machines: deadenylation by the Ccr4-not and Pan2-Pan3 complexes. Biochim. Biophys. Acta *1829*, 561–570.

Weill, L., Belloc, E., Bava, F.A., and Méndez, R. (2012). Translational control by changes in poly(A) tail length: recycling mRNAs. Nat. Struct. Mol. Biol. *19*, 577–585.

Yamashita, A., Chang, T.C., Yamashita, Y., Zhu, W., Zhong, Z., Chen, C.Y., and Shyu, A.B. (2005). Concerted action of poly(A) nucleases and decapping enzyme in mammalian mRNA turnover. Nat. Struct. Mol. Biol. *12*, 1054–1063.

Zeqiraj, E., Filippi, B.M., Goldie, S., Navratilova, I., Boudeau, J., Deak, M., Alessi, D.R., and van Aalten, D.M. (2009). ATP and MO25alpha regulate the conformational state of the STRADalpha pseudokinase and activation of the LKB1 tumour suppressor. PLoS Biol. *7*, e1000126, http://dx.doi.org/10.1371/journal.pbio.1000126.

Zheng, D., Ezzeddine, N., Chen, C.Y., Zhu, W., He, X., and Shyu, A.B. (2008). Deadenylation is prerequisite for P-body formation and mRNA decay in mammalian cells. J. Cell Biol. *182*, 89–101.

Molecular Cell, Volume *51*

Supplemental Information

Structure of the PAN3 Pseudokinase Reveals the Basis for Interactions with the PAN2 Deadenylase and the GW182 Proteins

Mary Christie, Andreas Boland, Eric Huntzinger, Oliver Weichenrieder, and Elisa Izaurralde

Inventory of Supplemental Information

Supplemental Experimental Procedures

Protein Expression and Purification

An *Nc* PAN3 cDNA encoding the PK and C-term domains (residues 234-656) was cloned into the pnEA-NpM expression vector (Diebold et al., 2011). *Dm* and *Hs* PAN3 PK+C fragments (residues 349-790 and 245-687, respectively) were cloned into the pnYC-NpM expression vector (Diebold et al., 2011). Proteins were expressed in *E. coli* BL21 (DE3) STAR cells in LB media, induced with 0.5 mM IPTG and grown overnight at 20ºC. Bacterial cells expressing MBP-*Dm* PAN3 were resuspended in 50 mM HEPES (pH 7.6), 250 mM NaCl, 5% (w/v) glycerol and 2 mM DTT supplemented with complete, EDTA-free protease inhibitor cocktail (Roche), lysozyme and DNase I. The cells were lysed by sonication at 4°C. Bacterial lysates were cleared by centrifugation and incubated with amylose resin (NEB) for 1 hr at 4°C. MBP-*Dm* PAN3 was eluted in resuspension buffer supplemented with 10 mM maltose. AMP-PNP and $MgCl_2$ (final concentration of 2 mM) were added to the protein solution, and the MBP tag was cleaved using HRV 3C protease diluted at 1:200 (protease-to-protein ratio) overnight at 4ºC. *Dm* PAN3 was further purified using gel filtration (HiLoad 26/60 Superdex 200; GE Healthcare) and ion exchange chromatography (HiTrap Q FF; GE Healthcare) in 10 mM HEPES (pH 7.6), 250 mM NaCl, 5% glycerol, 2 mM DTT and 2 mM $MgCl_2$. Purified *Dm* PAN3 protein was snap frozen in liquid nitrogen and stored at -80°C. *Nc* PAN3 and *H*s PAN3, PAN3 fragments and mutants (Tables S1 and S4) were subjected to similar purification procedures. Purified *Nc* PAN3 was concentrated to 5 mg/ml in 10 mM HEPES (pH 7.2), 200 mM NaCl and 1 mM DTT, snap frozen in liquid nitrogen and stored at -80°C. Purified *Hs* PAN3 was snap frozen in 10 mM Tris-HCl (pH 8.5), 250 mM NaCl, 5% glycerol and 2 mM DTT.

Crystallization

Crystals of wild-type *Nc* PAN3 and of the seleno-methionine-substituted *Nc* PAN3 were obtained by the sitting drop vapor diffusion method over a 50 µl reservoir at 22°C after mixing 0.2 µl protein solution with 0.2 µl reservoir solution containing 0.1 M MES (pH 6.5), 10% PEG 5000 and 12% isopropanol. Crystals were reproduced by hanging-drop vapor diffusion in 0.1 M HEPES (pH 7.5), 7% PEG 6000 and 6% isopropanol at 18°C, using 5 mg/ml *Nc* PAN3 protein solution containing 2 mM ATP-γ-S and 2 mM $MgCl_2$. Crystals of *Nc* $PAN3^M$ were obtained in 0.1 M HEPES (pH 7.0), 15% PEG monomethyl ether 5000 (MME) and 0.1 M potassium chloride, using 5 mg/ml protein solution containing 2 mM ATP- γ-S and 2 mM $MgCl_2$. Crystals were flash-frozen in liquid nitrogen in the reservoir solution supplemented with 25% glycerol as cryoprotectant. Crystals of *Dm* PAN3 were obtained by hanging drop vapor diffusion method at 22ºC in 220 mM Na-citrate (pH 7.0), HEPES (pH 6.8) and 2% PEG 3350, using 8 mg/ml *Dm* PAN3 protein supplemented with 2 mM AMP-PNP and 2 mM $MgCl_2$. Crystals were transiently soaked in mother liquor supplemented with 25% glycerol and flash-frozen in liquid nitrogen.

Data Collection and Structure Determination

Diffraction data were collected on a PILATUS 6M detector at the PXII beamline of the Swiss Light Source (SLS). Native data for *Dm* PAN3 and *Nc* PAN3 were collected at a wavelength of 1.0 Å. SAD data were collected at the SeMet peak (0.9792 Å) for the SeMet derivatized *Nc* PAN3 crystals. Diffraction images were processed with XDS (Kabsch, 2010). Se sites and phases were calculated using AutoSol (PHENIX package, Afonine et al., 2012). Iterative cycles of manual model

building in real space (COOT, Winn et al., 2011) and refinement directly against the SAD data (REFMAC; Vagin et al., 2004) were performed. Native crystals of *Nc* PAN3 wild-type, *Nc* PAN3^{M} and *Dm* PAN3 diffracted to a resolution of 3.3 Å, 2.8 Å and 3.6 Å, respectively, and molecular replacement was performed using PHASER (McCoy et al., 2007) using the SeMet *Nc* PAN3 as a search model. Final refinement rounds for the *Dm* PAN3 were done in REFMAC, keeping local NCS restraints and refining TLS parameters (one group per chain) in addition to individual B-factors. Final refinement rounds for the *Nc* PAN3 wild-type and mutant were done in PHENIX, keeping torsion NCS restraints and refining TLS parameters and individual B-factors. The structural images presented in figures were prepared with PyMOL (http://www.pymol.org). The refinement statistics are summarized in Table S3.

Multi-angle Static Laser-light Scattering (MALLS)

Purified *Dm* PAN3, *Nc* PAN3 and *Hs* PAN3 proteins were loaded onto a Superdex-200 10/300 GL column (GE Healthcare) equilibrated with a buffer containing 10 mM HEPES (pH 7.6), 250 mM NaCl, 5% glycerol and 2 mM DTT. The column was connected to a miniDAWN TREOS multi-angle laser light scattering (MALLS) detector and Optilab rEX differential refractometer (Wyatt Technologies). Molecular weight calculations were performed using ASTRA software (Wyatt Technologies).

Thermal Shift Assays

Thermal shift assays were performed as previously described (Cummings et al., 2006). Sypro orange (5x final concentration, Sigma) was added to purified PAN3 proteins (diluted to ~ 2 μM in 50 mM HEPES [pH 7.6], 50 mM HEPES [pH 7.2] and 50 mM Tris-HCl [pH 8.5] for *Dm* PAN3, *Nc* PAN3 and *Hs* PAN3, respectively) in the

presence or absence of 2 mM of nucleotide (resuspended in 50 mM HEPES [pH 7.6]) and 2 mM $MgCl_2$. Protein-dye solutions were dispensed into the wells of a 96-well RT-PCR plate (Thermo Scientific) and loaded into a Bio-Rad C1000 thermal cycler. The temperature was gradually increased from 10°C to 95°C in 0.5°C increments in 10 sec steps. The fluorescence of the SYPRO orange dye was followed at 570 nm as a function of temperature. The data were analyzed using CFX Manager software (Bio-Rad) and apparent melting temperatures were obtained via the maximum value of the first derivative of the melting curve. Assays were performed in triplicate, and the apparent melting temperature (Tm_{app}) represent the mean value obtained.

GST Pull-Down Assays

GST-tagged TNRC6 silencing domains were expressed in *E. coli* BL21(DE) STAR cells following induction with 0.5 mM IPTG at 20ºC overnight. Cells were resuspended in 50 mM Tris-HCl [pH 8.5], 250 mM NaCl, 5% glycerol and 2 mM DTT and lysed by sonication at 4ºC. Cleared lysates of cells expressing GST-TNRC6-SDs (6C-SD or 6B-SD) and MBP-PAN3 were mixed and incubated with Protino Glutathione agarose 4B beads (Machery Nagel) for 1 hr at 4°C. The beads were washed three times in resuspension buffer and eluted with SDS-PAGE loading buffer. Bound proteins were analyzed using 12% SDS-PAGE.

Coimmunoprecipitation Assays

Plasmids for the expression of PAN3, PAN2 and GW182/TNRC6s proteins in human and *Dm* S2 cells were described previously (Braun et al., 2011; Huntzinger et al., 2013). PAN3 mutants were generated by site-directed mutagenesis using the QuikChange Site-Directed Mutagenesis kit from Stratagene and the appropriate

oligonucleotide sequences. For coimmunoprecipitation assays in human cells, HEK293 cells were grown in 10-cm dishes and transfected using Lipofectamine 2000 (Invitrogen). *Dm* S2 cells were grown in 6-well dishes and transfected using Effectene (Qiagen) transfection reagent and harvested 3 days after transfection. Immunoprecipitation was performed as described by Braun *et al*. (2011). Cell lysates were treated with Micrococcal nuclease for 30 min and spun at 18,000 x *g* for 15 min at 4°C prior to immunoprecipitation.

HA- and GFP-tagged proteins were detected using HRP-conjugated monoclonal anti-HA (Roche 3F10; 1:5,000) and anti-GFP antibodies (Roche 11814460001; 1:2,000), respectively. V5-tagged proteins were detected with anti-V5 antibodies (Invitrogen, 1:5,000). All western blots were developed with the ECL western blotting detection system (GE Healthcare) as recommended by the manufacturer.

Tethering Assays in S2 Cells

The tethering assays in S2 cells were performed essentially as described previously (Braun et al., 2011). S2 cells were grown in 6-well dishes and transfected using Effectene (Qiagen). The following plasmids were cotransfected: 0.1 μg reporter plasmid (F-Luc-5BoxB or F-Luc), 0.4 μg pAc5.1-R-Luc as transfection control and 25 ng of pAc5.1λN-HA construct for the expression of λN-HA-PAN3 (wild type or mutants). R-Luc and F-Luc activities were measured 3 days after transfection using the Dual-Luciferase Reporter Assay System (Promega). To measure the mRNA half-life, transfected cells were treated with actinomycin D (5 μg/ml final concentration) 3 days after transfection and harvested at the indicated time points. Total RNA was isolated using Trifast Reagent (Peqlab) and analyzed by Northern blot as described

previously (Braun et al., 2011).

ATPase Assay

For the experiment shown in Figure S4A, *Dm* PAN3 (65 pmol) or calf intestinal alkaline phosphatase (20 fmol) was incubated at 25°C for 20 hours with 10 nM [γ-^{32}P]ATP (1μM cold ATP), in a total reaction volume of 20 μl containing buffer A (75 mM Bis-Tris propane [pH 7.2], 100 mM KCl, 150 mM NaCl, 1 mM DTT, 10 mM $MgCl_2$ and 1 mM DTT). The reaction was stopped by adding 1 volume 0.5 M EDTA and 1.5 μl aliquots were spotted onto polyethyleneimine cellulose plates for chromatography (Merck) that were developed in 0.75 M LiCl. ATPase activity was assayed by the formation of [^{32}P]P_i.

REFERENCES

Afonine, P.V., Grosse-Kunstleve, R.W., Echols, N., Headd, J.J., Moriarty, N.W., Mustyakimov, M., Terwilliger, T.C., Urzhumtsev, A., Zwart, P.H., and Adams, P.D. (2012) Towards automated crystallographic structure refinement with phenix.refine. Acta Crystallogr. D Biol. Crystallogr. *68*(Pt 4), 352–367.

Cummings, M.D., Farnum, M.A., and Nelen, M.I. (2006) Universal screening methods and applications of ThermoFluor. J. Biomol. Screen *11*, 854–863.

Diebold, M.L., Fribourg, S., Koch, M., Metzger, T., and Romier, C. (2011) Deciphering correct strategies for multiprotein complex assembly by co-expression: application to complexes as large as the histone octamer. J. Struct. Biol. *175*, 178–188.

Kabsch, W. (2010) XDS. Acta Crystallogr. D Biol. Crystallogr. *66*, 125–132.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. J. Appl. Cryst. *40*, 658–674.

Vagin, A.A., Steiner, R.A., Lebedev, A.A., Potterton, L., McNicholas, S., Long, F., Murshudov, G.N. (2004). REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. Acta Crystallogr. D Biol. Crystallogr. *60* (Pt 12 Pt 1), 2184–2195.

Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G., McCoy, A., et al. (2011). Overview of the CCP4 suite and current developments. Acta Crystallogr. D Biol. Crystallogr. 67(Pt 4), 235–242.

SUPPLEMENTAL FIGURES

Figure S1

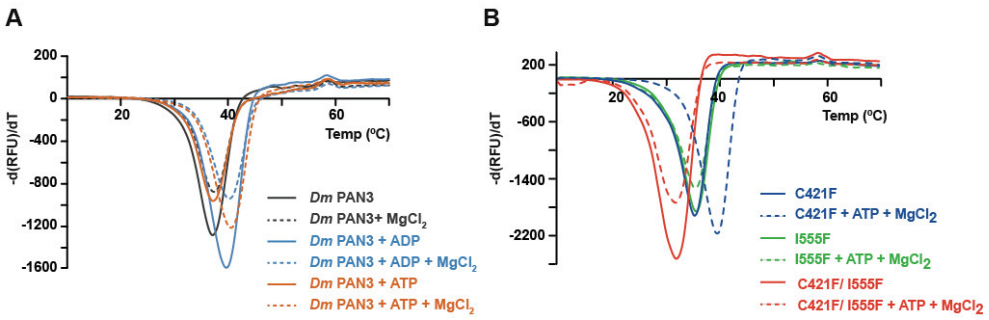


Figure S1. The *Dm* PAN3 PK domain binds ATP analogs. Related to Figure 2.

(A and B) Thermal-shift assays performed with the *Dm* PAN3 PK+C-term fragment (wild-type or mutants) in the presence and absence of nucleotide and Mg^{2+}. The apparent melting temperatures are indicated in Tables S2 and S5.

Figure S2

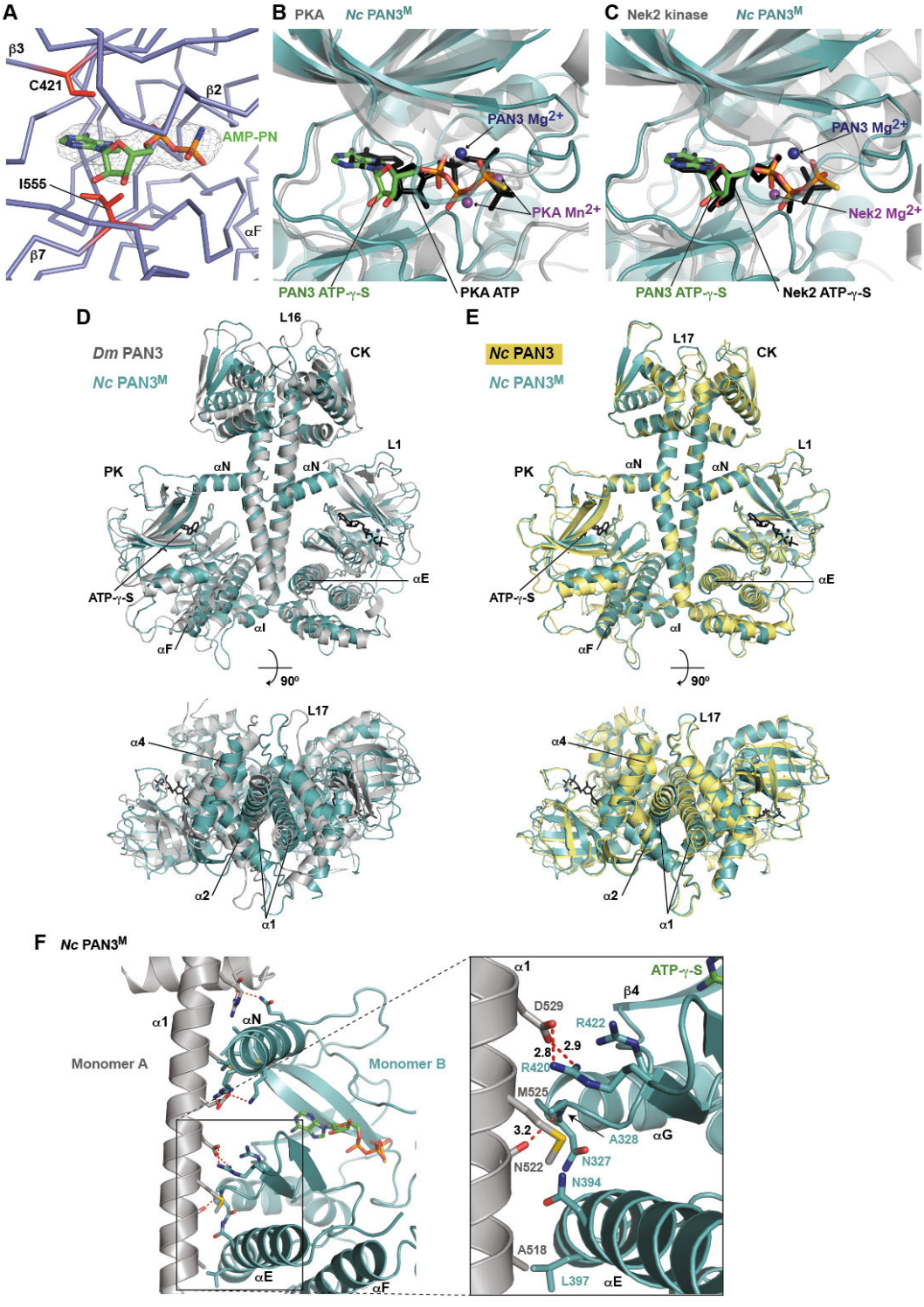

A
β3
C421
β2
AMP-PN
I555
αF
β7
B PKA Nc PAN3M
PAN3 Mg2+
PKA Mn2+
PAN3 ATP-γ-S
PKA ATP
C Nek2 kinase Nc PAN3M
PAN3 Mg2+
Nek2 Mg2+
PAN3 ATP-γ-S
Nek2 ATP-γ-S
D
Dm PAN3
Nc PAN3M
L16
CK
L1
PK
αN
αN
ATP-γ-S
αE
αI
αF
90°
L17
α4
α2
α1
E
Nc PAN3
Nc PAN3M
L17
CK
L1
PK
αN
αN
ATP-γ-S
αE
αI
αF
90°
L17
α4
α2
α1
F Nc PAN3M
α1
αN
Monomer A
Monomer B
αE
αF
α1
ATP-γ-S
β4
D529
R422
2.8
2.9
R420
M525
3.2
A328
αG
N522
N327
N394
A518
L397
αE

Figure S2. **Structures of the *Dm* and *Nc* PAN3 PK+C-term domain. Related to Figures 2 and 3.**

(A) Close-up view of the ATP-binding pocket of *Dm* PAN3 with simulated annealing Fo-Fc omit map contoured at 3σ shown in grey mesh. The AMP-PN molecule is overlaid and the residues that prevent ATP binding when mutated are shown in red stick representation.

(B and C) Close-up view of the ATP-binding pocket of *Nc* $PAN3^{M}$ superimposed onto the ATP-binding site of the canonical PKA ((B); PDB ID 1ATP) or Nek2 kinase ((C); PDB ID 2W5B). Strands β2-β5 of *Nc* $PAN3^{M}$ were used for the superposition. The *Nc* $PAN3^{M}$ ATP-γ-S and Mg^{2+} molecules are shown in green stick and blue sphere representation respectively. The ATP and Mn^{2+} molecules from PKA are shown in black stick and purple sphere representation, respectively. The ATP-γ-S and Mg^{2+} molecules from the Nek2 kinase are shown in black stick and purple sphere representation, respectively. For clarity, the PKA and Nek2 kinase structures are shown in transparent grey cartoon representation.

(D) Superposition of the wild-type *Dm* PAN3 (grey) and *Nc* $PAN3^{M}$ structures (teal) in two orientations.

(E) Superposition of the *Nc* PAN3 wild-type (yellow) and mutant (teal) in two orientations.

(F) Close-up view of the *Nc* $PAN3^{M}$ mutant dimerization interface. Monomer A and B are shown in gray and teal, respectively. Atomic distances are indicated in Å.

Figure S3

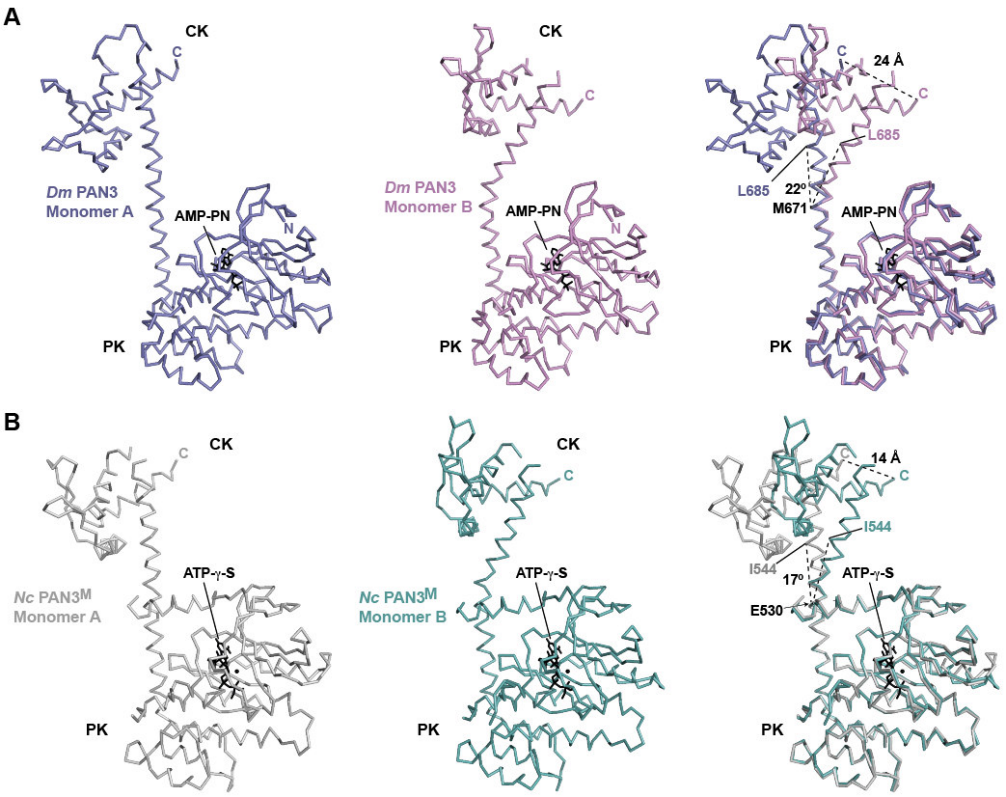


Figure S3. **Comparison of monomeric *Dm* and *Nc* PAN3 PK and C-term domains. Related to Figure 2.**

(A) Ribbon representation of the individual monomers and superposition of both monomers of *Dm* PAN3. Residues 378-678 were used for superposition. The different angles between the CK domains relative to the PK domain caused by the helical kink are indicated.

(B) Ribbon representation of the individual monomers and superposition of both monomers of *Nc* PAN3^{M}. Residues 235-537 were used for superposition. The different angles between the CK domains relative to the PK domain caused by the helical kink are indicated.

Figure S4

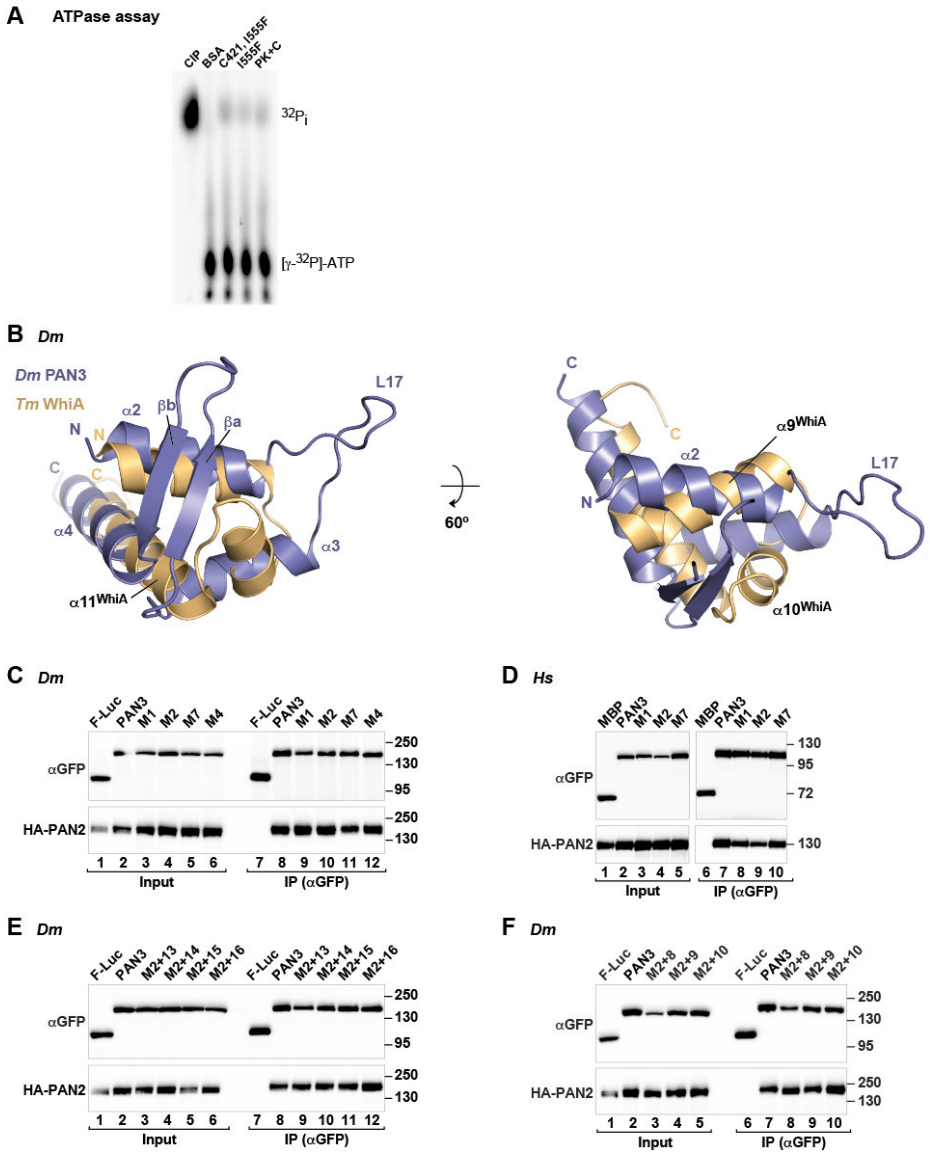


Figure S4. The PAN3 PK domain does not have ATPase activity *in vitro*. Related to Figures 3, 4 and 5.

(A) Thin-layer chromatography analysis of *Dm* PAN3 (wild type and mutants) ATPase activity. Calf intestinal alkaline phosphatase served as the positive control. The positions of the [γ-^{32}P] phosphate-labeled ATP and free [^{32}P] released phosphate are indicated.

(B) Superposition of the *Dm* PAN3 CK domain (blue) and *Tm* WhiA HTH domain structures (orange; PDB ID 3HYI). Secondary structure elements are labeled.

(C–F) The interaction of GFP-PAN3 (wild type or mutants) with HA-PAN2 was analyzed by coimmunoprecipitation in *Dm* S2 and *Hs* HEK293 cells as described in Figure 1.

Figure S5

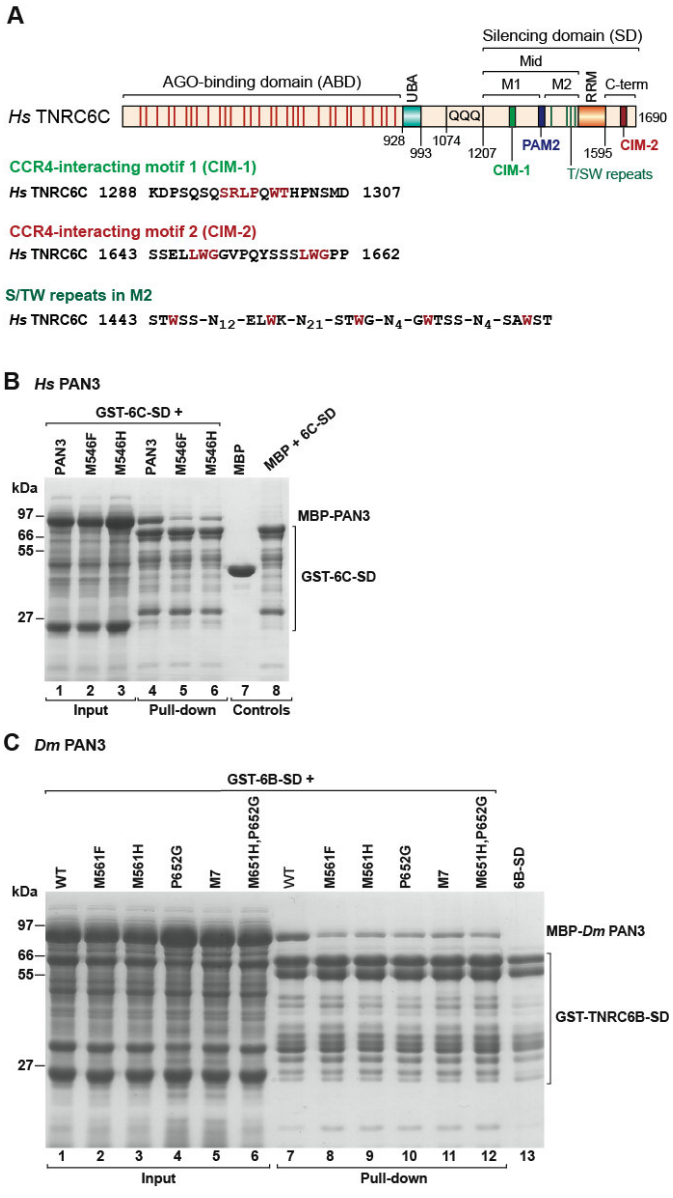


Figure S5. The PAN3 W-binding pocket is required for binding to the TNRC6 proteins *in vitro*. Related to Figure 6.

(A) Domain organization of *Hs* TNRC6C. ABD: AGO-binding domain; UBA: ubiquitin associated-like domain; QQQ: region rich in glutamine; Mid: middle region containing the PAM2 motif (dark blue), which divides the Mid region into the M1 and M2 regions; RRM: RNA recognition motif; C-term: C-terminal region and SD: silencing domain. The position of the conserved CIM-1, CIM-2 motifs are indicated. Amino acid positions at domain boundaries are indicated below the protein outlines. Vertical red lines within the ABD indicate the positions of GW repeats involved in AGO-bonding. Vertical green lines within the M2 region indicate the positions of T/SW repeats implicated in PAN3-binding. The sequences of the CIM-1, CIM-2 and T/SW motifs are indicated. Amino acids substituted by Alanine are shown in in red.

(B) Interaction between recombinant MBP-*Hs* PAN3 (PK+C domain wild type or mutants) and GST-6C-SD. Inputs (5%) and bound fractions (20%) were analyzed by SDS-PAGE. MBP served as a negative control.

(C) Interaction between recombinant MBP-*Dm* PAN3 (PK+C domain wild type or mutants) and GST-6B-SD analyzed as described in (A).

Figure S6

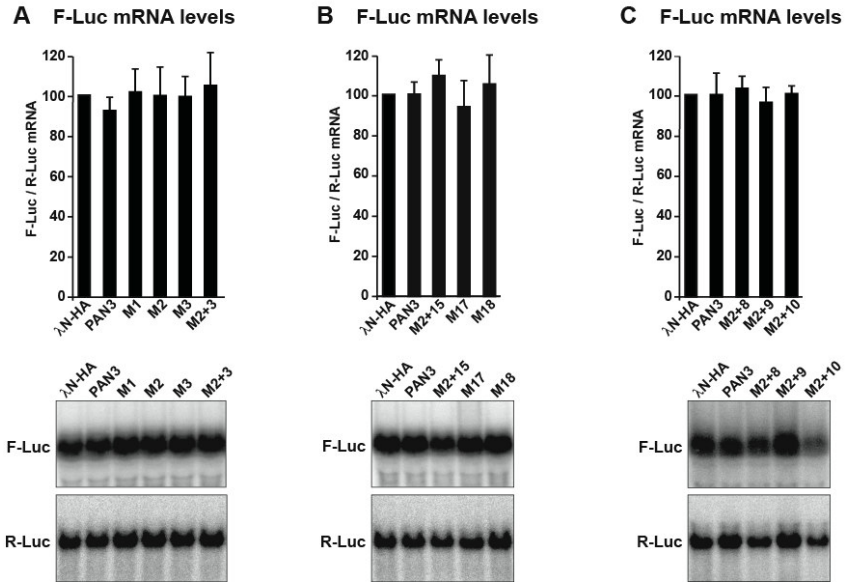


Figure S6. The integrity of the nucleotide binding site and PAN2-binding are required for mRNA degradation. Related to Figure 7.

(A, B and C) A F-Luc reporter lacking the BoxB hairpins was coexpressed with the indicated proteins. The corresponding experiment with a F-Luc-5BoxB reporter is shown in Figure 7. The levels of the F-Luc mRNA were normalized to that of R-Luc mRNA. The values were set at 100% in cells expressing the λN-HA peptide (white bar). Bars represent mean values and error bars represent standard deviations from at least three independent experiments. Representative Northern blots are shown below the graphs.

Table S1. Multi-angle laser light scattering analysis of PAN3 proteins

Related to Figures 1, 3, 4 and 6.

Protein construct	**Alternative names**	**Expected molecular weight (kDa)**	**Observed molecular weight (kDa)**
Dm PAN3 349-790	*Dm* PAN3 PK+C	50.9	100
Dm PAN3 349-790 + AMP-PNP + $MgCl_2$	*Dm* PAN3 PK+C	50.9	96
Dm PAN3 349-655	*Dm* PAN3 PK	35.2	38
Dm PAN3 641-790	C-term + αI	18.1	18
Dm PAN3 PK+C CI555F	ATP binding mutant	50.9	97
Dm PAN3 PK+C C421F	ATP binding mutant	50.9	96
Dm PAN3 PK+C C421F,I555F	ATP binding mutant	51.0	100
Dm PAN3 PK+C F658E,L662E,L665E	**M1** dimerization mutant	51.0	53
Dm PAN3 PK+C L676E,G683E	**M2** dimerization mutant	51.0	51
Nc PAN3 234-656	*Nc* PAN3 PK+C	49.2	104
Hs PAN3 245-687	*Hs* PAN3 PK+C	50.9	100
Hs PAN3 PK+C M546F	*Hs* PAN3 PK+C W-binding mutant	50.9	102
Hs PAN3 PK+C M546H	*Hs* PAN3 PK+C W-binding mutant	50.9	101
Hs PAN3 PK+C M546F,P547G	**M7** W-binding mutant	50.9	100
Hs PAN3 PK+C F553E,L557E,A560E	**M1** dimerization mutant	51.0	56
Hs PAN3 PK+C L571E,G578E	**M2** dimerization mutant	51.0	86

Table S2. Apparent melting temperatures of *Nc*, *Dm* and *Hs* PAN3 PK+C fragments. Related to Figures 2 and 3.

Condition	Tm_{App} *Nc* PAN3	Tm_{App} *Dm* PAN3	Tm_{App} *Hs* PAN3
Buffer only	45.2	37.0	42.5
2 mM $MgCl_2$	45.5	37.2	42.3
2 mM AMP	52.5	40.2	47.5
2 mM ADP	53.7	40.5	46.0
2 mM ATP	45.0	37.2	43.5
2 mM AMP + $MgCl_2$	52.5	41.0	47.2
2 mM ADP + $MgCl_2$	52.5	39.5	47.2
2 mM ATP + $MgCl_2$	54.7	40.7	51.3
2 mM AMPPNP + $MgCl_2$	54.3	39.5	49.5
2 mM ATP-γ-S + $MgCl_2$	54.5	39.5	50.5
2 mM GMP + $MgCl_2$	45.0	37.5	42.3
2 mM GDP + $MgCl_2$	44.8	37.2	42.0
2 mM GTP + $MgCl_2$	45.0	37.2	42.0

Table S3. Mutants used in this study

Related to Figures 2, 4, 6 and 7, and Table 1.

Name	***Dm* PAN3** (NP_647767.1)	***Hs* PAN3** (BC128180.1)	**Location**
M1	F658E,L662E,L665E	F553E,L557E,A560E	Dimer interface
M2	L676E,G683E	L571E,G578E	Dimer interface
M3	T408F,C421F,I555F	C319F,L447F	Nucleotide binding
M2+M3	T408F,C421F,I555F+M2	C319F,L447F+M2	Nucleotide binding
M7	M651F,P652G	M546F,P547G	W-pocket
M8	Δ460-465 replaced by GSSG	Δ358-363 replaced by GSSG	PK surface
M9	R429D	R327D	PK surface
M10	D575R	D468R	PK surface
M13	S708A,T710A	S603A,T605A	CK
M14	K717S,D721S,H725S	-	CK
M15	H739E,K746E,D748K	H634E,K641E,D643K	CK
M16	R760T,D761T	R655T,D656T	CK
M17	M2+M15+M16		CK
M18	M2+M14+M15+M16		CK

The *Nc* PAN3^{M} mutant carries the following substitutions: L510M, G511P, G512M, T514G, T515A, H516R, L517F. The *Nc* PAN3 residues were substituted with residues found at equivalent positions in *Dm* PAN3 with the aim to generate a W-binding pocket and promote alternative crystal packing arrangements for the *Nc* PAN3 protein. However, the pocket formed in *Nc* PAN3^{M} protein appears to be too narrow to accommodate a tryptophan.

TNRC6C mutants CIM1+CIM2 and M2 are as shown in Figure S5A.

Table S4. Apparent melting temperatures of PAN3 mutants

Related to Figures 3, 4 and 6, and Table 1.

	***Dm* PAN3 C421F**	***Dm* PAN3 I555F**	***Dm* PAN3 I555F/ C421F**	***Dm* PAN3 M1**	***Dm* PAN3 M2**	***Hs* PAN3 M1**	***Hs* PAN3 M2**	***Hs* PAN3 M7**	***Nc* PAN3^{M}**
Buffer only	35.5	35.7	32.0	36.0	36.5	32.5	34.0	38.5	44.8
2 mM $MgCl_2$	35.5	35.7	32.0	36.0	35.8	32.7	33.7	38.8	44.5
2 mM ADP + $MgCl_2$	37.5	35.7	32.0	37.3	37.5	37.3	37.5	40.33	50.8
2 mM ATP + $MgCl_2$	39.8	35.7	32.0	39.5	39.3	41.3	39.5	43.67	56.0

nature structural & molecular biology

Structure and assembly of the NOT module of the human CCR4–NOT complex

Andreas Boland[1,2], Ying Chen[1,2], Tobias Raisch[1,2], Stefanie Jonas[1,2], Duygu Kuzuoğlu-Öztürk[1], Lara Wohlbold[1], Oliver Weichenrieder[1] & Elisa Izaurralde[1]

The CCR4–NOT deadenylase complex is a master regulator of translation and mRNA stability. Its NOT module orchestrates recruitment of the catalytic subunits to target mRNAs. We report the crystal structure of the human NOT module formed by the CNOT1, CNOT2 and CNOT3 C-terminal (-C) regions. CNOT1-C provides a rigid scaffold consisting of two perpendicular stacks of HEAT-like repeats. CNOT2-C and CNOT3-C heterodimerize through their SH3-like NOT-box domains. The heterodimer is stabilized and tightly anchored to the surface of CNOT1 through an unexpected intertwined arrangement of peptide regions lacking defined secondary structure. These assembly peptides mold onto their respective binding surfaces and form extensive interfaces. Mutagenesis of individual interfaces and perturbation of endogenous protein ratios cause defects in complex assembly and mRNA decay. Our studies provide a structural framework for understanding the recruitment of the CCR4–NOT complex to mRNA targets.

The CCR4–NOT deadenylase complex has a crucial role in post-transcriptional mRNA regulation. It catalyzes the removal of mRNA poly(A) tails and consequently represses translation and promotes mRNA degradation[1,2]. Remarkably, the CCR4–NOT complex can also repress translation independently of deadenylation, and it facilitates dissociation of the cytoplasmic poly(A)-binding protein (PABP) from mRNA poly(A) tails[3–6].

In addition to its central role in bulk and specific mRNA degradation, the CCR4–NOT complex has been implicated in transcription initiation and elongation, in DNA repair and in ubiquitination and protein modification[1,2]. Given the diverse activities associated with the CCR4–NOT complex, it is not surprising that it has a role in a wide range of biological processes, including cell proliferation, apoptosis, oogenesis and embryogenesis, spermatogenesis, heart function, bone formation and energy metabolism[1,2].

The conserved core of the CCR4–NOT complex consists of two major modules: a catalytic module comprising two deadenylases (CAF1 or its paralog POP2, and CCR4a or its paralog CCR4b) and the NOT module, which minimally consists of NOT1, NOT2 and NOT3. Additional subunits within the complex have been described, including NOT4, CAF40 and the species-specific subunits CAF130, NOT10 and NOT11 (refs. 1,2). NOT1 functions as a modular scaffold to provide binding sites for the NOT10–NOT11 module at its N terminus, the CAF1–CCR4 catalytic module and CAF40 in its middle region and the NOT2 and NOT3 subunits at the C terminus. Thus, NOT1 is essential for the assembly of the complete CCR4–NOT complex[3,7–16].

The precise molecular function of the NOT module remains unclear. Current evidence indicates that it regulates the stability and activity of the catalytic module. Another crucial role for the NOT module is the recruitment of the CCR4–NOT complex to a plethora of specific mRNAs[2,17]. These include microRNA targets and mRNAs containing AU-rich elements, to which the CCR4–NOT complex is recruited through interactions of NOT1 with GW182 proteins and tristetraprolin, respectively[4,18–21]. Furthermore, a multitude of translational regulators, which include Nanos, Bicaudal-C, CUP and Smaug, recruit the CCR4–NOT complex to their targets through interactions with NOT-module subunits[17].

The mechanistic understanding of the assembly, regulation and function of the NOT module requires elucidation of its three-dimensional structure, specifically of the C-terminal regions of NOT1, NOT2 and NOT3, which mediate NOT-module assembly[1–3,12,22,23]. These regions comprise a highly conserved NOT1 superfamily homology (SH) domain in NOT1 and a conserved NOT-box domain[22,23] in NOT2 and NOT3. Currently there are no structural models available for these regions, owing to the lack of sequence homology to known structural folds.

To provide the missing structural framework for understanding the assembly and functions of the NOT module, we determined the crystal structure of a ternary complex formed by the human CNOT1-C, CNOT2-C and CNOT3-C regions. This was achieved by a stepwise approach wherein we first determined the boundaries and crystal structures of the isolated human CNOT2 and CNOT3 NOT-box domains and of the *Chaetomium thermophilum* (*Ct*) NOT1 SH domain. We then used this information for the assembly and structure determination of the ternary CNOT1–CNOT2–CNOT3 complex. The structures reveal a rigid scaffold for the NOT1 SH domain,

[1]Department of Biochemistry, Max Planck Institute for Developmental Biology, Tübingen, Germany. [2]These authors contributed equally to this work. Correspondence should be addressed to E.I. (elisa.izaurralde@tuebingen.mpg.de) or O.W. (oliver.weichenrieder@tuebingen.mpg.de).

Received 30 May; accepted 20 August; published online 13 October 2013; doi:10.1038/nsmb.2681

consisting of two perpendicularly arranged stacks of α-helices, and an SH3-like fold for the CNOT2 and CNOT3 NOT boxes, which mediate heterodimerization. Perhaps the most remarkable structural features are the intertwined N-terminal extensions of CNOT2 and CNOT3 that wrap around their partners in the heterodimer and also form an extended interface with CNOT1. Functional studies reveal the relevance of the interaction interfaces for complex assembly and mRNA degradation.

RESULTS

Mutual interactions between CNOT1, CNOT2 and CNOT3

Our previous studies indicated that the interactions between *Drosophila melanogaster* (*Dm*) NOT1, NOT2 and NOT3 are mediated by their C-terminal regions[3]. Coimmunoprecipitations in human cells demonstrated that this mode of interaction is conserved in the human orthologs (**Supplementary Fig. 1a–e**). Moreover, although CNOT1 interacted with both CNOT2 and CNOT3, the interaction with CNOT3 was enhanced in cells coexpressing CNOT2 (**Supplementary Fig. 1f**).

To determine whether the observed interactions are direct and to further define the domain boundaries, we performed pulldown assays with isolated C-terminal protein fragments expressed in *Escherichia coli*. We observed direct interactions for all possible binary domain combinations (**Supplementary Fig. 1g–j**). Further analysis indicated that the CNOT2-C fragment (residues 344–540) interacts with both CNOT1 and CNOT3, whereas a shorter fragment (residues 429–540) binds CNOT3 but is not sufficient for CNOT1 binding (**Fig. 1a** and **Supplementary Fig. 1g,h**). The smallest CNOT3 fragment that maintained CNOT2 binding comprised residues 656–753 (**Supplementary Fig. 1i**). On the basis of these results and our structural analysis, we will refer to residues 429–525 of CNOT2 and residues 656–747 of CNOT3 as the NOT-box domains (**Fig. 1a**). We conclude that the NOT-box domains mediate the CNOT2-CNOT3 interaction, whereas additional N-terminal sequences are required for CNOT1 binding.

The NOT1 SH domain features two perpendicular HEAT stacks

To determine the structure of the CNOT1-C domain, we initially identified a proteolytically stable fragment in CNOT1 (residues 1842–2371; **Fig. 1a**) that was expressed in a soluble form but failed to crystallize. We then expressed an equivalent fragment of *Ct* NOT1 (residues 1676–2193; **Supplementary Fig. 2**), which yielded crystals that diffracted X-rays to 3.2-Å resolution (**Table 1**).

The overall architecture of *Ct* NOT1-C is L-shaped and consists of two tightly interacting subdomains connected by a surface-attached and well-structured linker loop (loop L11; **Fig. 1b–d**). The two subdomains correspond to two perpendicularly arranged stacks of α-helices: an N-terminal subdomain (N-SD, helices α1–α14) that runs left to right and a C-terminal subdomain (C-SD, helices α15–α23) that runs bottom to top (**Fig. 1b–d**). The helices are packed as antiparallel HEAT-like pairs, with six hairpins in the N-SD and four hairpins in the C-SD. The arrangement of HEAT-like repeats in perpendicular stacks is different from the usually continuous arrangement[24], and it occurs more than once in the context of NOT1 (ref. 10).

Our structural analysis redefines the boundaries of the NOT1 SH domain and reveals that most of the highly conserved residues map to the extensive interface (1,151 Å^2) between the N- and C-SDs (**Supplementary Fig. 2**), indicating that the mutual arrangement

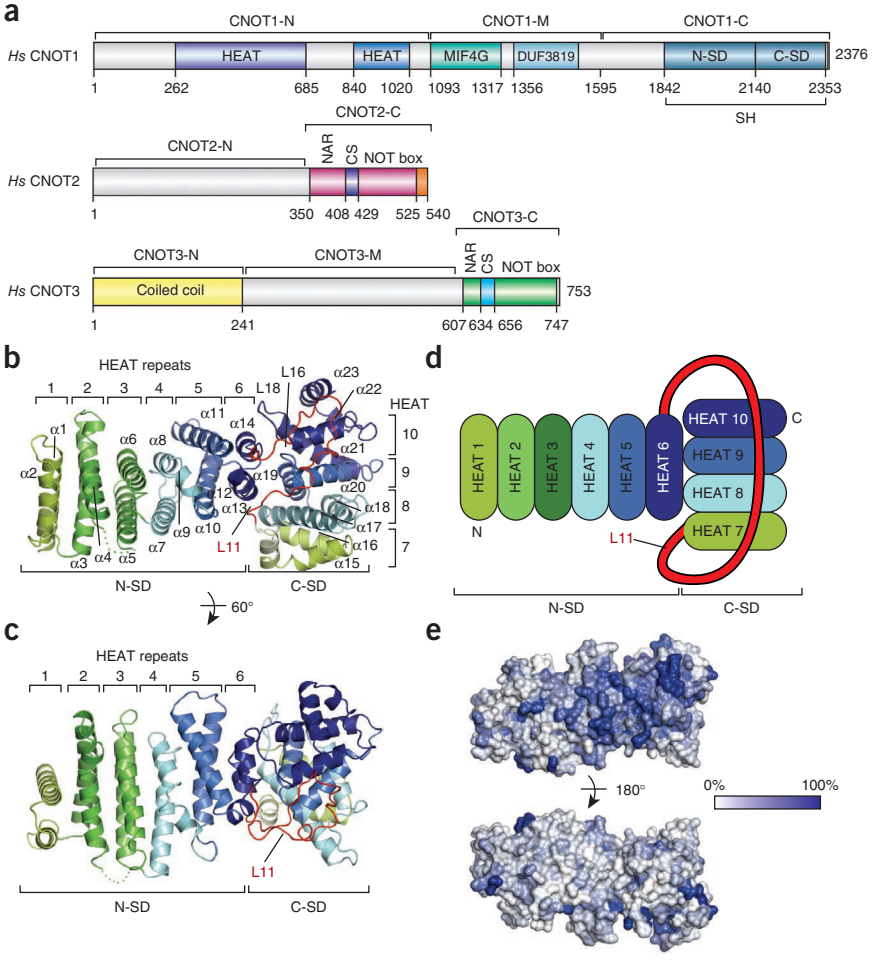


Figure 1 Domain organization of *Homo sapiens* (*Hs*) CNOT1–CNOT3 and the structure of the *Ct* NOT1 superfamily homology domain. (**a**) Diagrams of CNOT1 with N-terminal, middle and C-terminal regions (CNOT1-N, CNOT1-M and CNOT1-C, respectively). The CNOT1-N consists of two HEAT-repeat domains. CNOT1-M contains a MIF4G domain and a domain of unknown function (DUF3819). CNOT1-C contains the NOT1 superfamily homology (SH) domain. The C-terminal regions of both CNOT2 and CNOT3 contain a NOT1 anchor region (NAR), a connector sequence (CS) and a NOT-box domain (38% sequence identity for the CNOT2 and CNOT3 NOT boxes)[22,23]. CNOT3 displays an N-terminal coiled coil domain (CNOT3-N) and a linker region (CNOT3-M). The numbers below the protein outline represent amino acid positions at the domain boundaries as defined in this study. (**b**,**c**) Cartoon representations of the *Ct* SH domain. Two views, related by 60° rotation, are shown in **b** and **c**. The N- and C-terminal subdomains (N-SD and C-SD, respectively) are each colored in a gradient from green to dark blue from the N to C terminus. HEAT repeats are indicated. Loop L11 linking the two lobes is highlighted in red. (**d**) Schematic drawing of the *Ct* NOT1 structure with the same colors as in **b**. (**e**) Surface conservation of NOT1 SH domain colored in a gradient from white to blue. The orientation of the top structure corresponds to that shown in **c**.

Table 1 Data collection and refinement statistics

	Ct NOT1 (Se)	*Ct* NOT1	*Hs* CNOT2 (Se)	*Hs* CNOT2	*Hs* CNOT3 (Se)	*Hs* CNOT3	*Hs* CNOT1–2–3
Data collection[a]							
Space group	$P2_12_12_1$	$P2_1$	$P4_1$	$P6_522$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12$
Cell dimensions							
a, *b*, *c* (Å)	126.6, 129.6, 154.1	77.8, 127.2, 130.2	76.4, 76.4, 86.3	64.8, 64.8, 412.3	52.2, 97.3, 142.3	52.3, 97.6, 141.5	91.6, 165.9, 78.8
α, β, γ (°)	90, 90, 90	90, 93.1, 90	90, 90, 90	90, 90, 120	90, 90, 90	90, 90, 90	90, 90, 90
Resolution (Å)	90.9–3.6 (3.69–3.6)	90.9–3.2 (3.28–3.2)	76.4–3.4 (3.49–3.4)	49.3–2.4 (2.46–2.4)	80.3–2.5 (2.57–2.5)	49.1–2.4 (2.46–2.4)	48.5–3.2 (3.28–3.2)
R_{sym} (%)	17.2 (73.1)	10.6 (54.5)	16.2 (68.4)	9.6 (80.2)	8.6 (73.5)	5.4 (70.4)	11.9 (55.9)
$I / \sigma I$	16.4 (5.1)	11.3 (2.9)	17.5 (4.9)	13.2 (2.4)	17.7 (3.7)	18.1 (3.2)	10.1 (2.2)
Completeness (%)	100.0 (100.0)	99.9 (99.9)	99.7 (96.0)	99.2 (99.1)	99.9 (99.1)	99.5 (99.8)	98.9 (98.9)
Redundancy	26.8 (27.9)	5.4 (5.2)	20.1 (18.1)	6.7 (6.7)	12.9 (12.9)	5.9 (6.2)	3.3 (3.4)
Refinement							
Resolution (Å)		63.6–3.2		49.3–2.4		49.1–2.4	48.5–3.2
No. reflections		41,820		21,287		28,937	20,241
R_{work} / R_{free} (%)		21.7 / 25.9		21.7 / 26.3		22.1 / 25.7	22.4 / 27.3
No. atoms							
Protein		15,737		3,483		4,984	6,864
Water		–		102		76	–
B factors (Å^2)							
Protein		88		51		69	71
Water		–		48		52	–
r.m.s. deviations							
Bond lengths (Å)		0.002		0.002		0.002	0.002
Bond angles (°)		0.53		0.64		0.62	0.52

[a]One crystal was used per data set. Values in parentheses are for highest-resolution shell.

between the two subdomains is evolutionarily conserved. Additional conserved surface residues map to only one of the two helical surfaces of the domain (**Fig. 1e**).

The NOT-box domain adopts an SH3-related fold

The CNOT3 NOT-box domain crystallized as a homodimer (2.4-Å resolution; **Table 1** and **Fig. 2a**), consistent with the results from multiangle static laser light scattering and size-exclusion chromatography, which also indicated that the protein was dimeric in solution (**Supplementary Fig. 3a**).

The structure of the NOT-box domain consists of three N-terminal α-helices (α1–α3) followed by a C-terminal SH3-like five-stranded open β-barrel (β1–β5), with β-strands β4 and β5 strongly bent (**Fig. 2a**,**b**). The N-terminal helices form an antiparallel bundle that packs against one side of the β-barrel. Notably, the inner core of the β-barrel and the interface with the N-terminal helices consist almost exclusively of aromatic side chains (**Supplementary Fig. 3b**,**c**).

The N-terminal helices also provide the interface for the dimerization of the NOT-box domain (**Fig. 2a**). The dimer has a two-fold symmetry with a parallel arrangement of the α2 helices, which provide

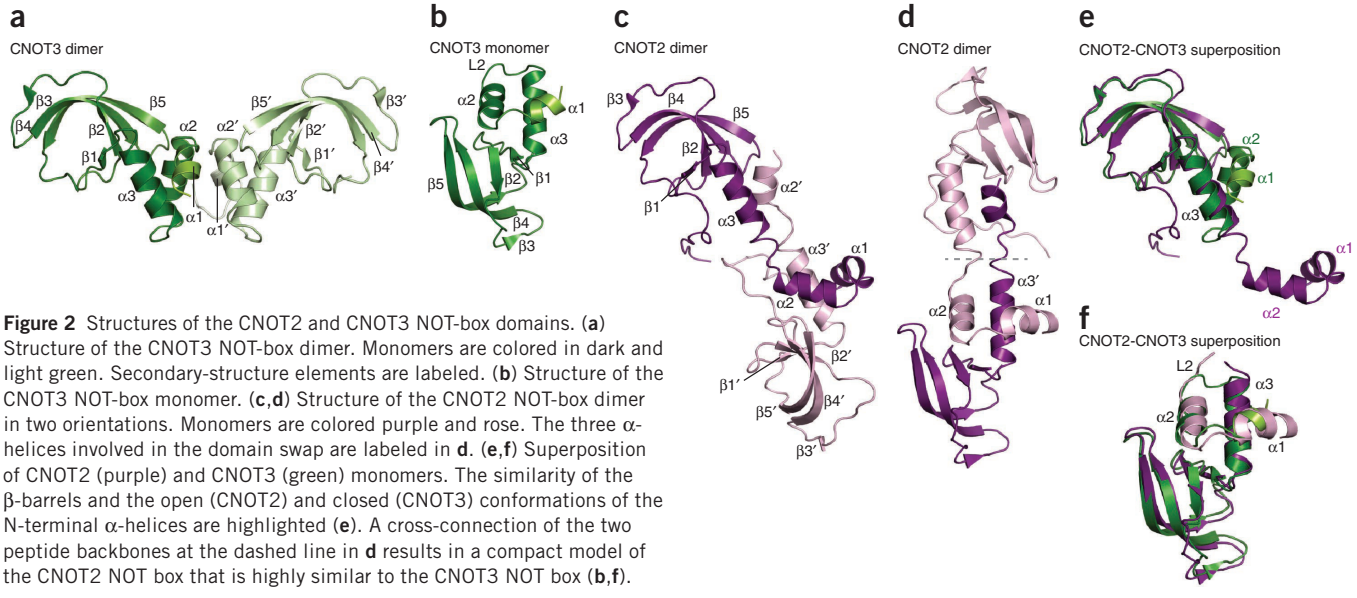


Figure 2 Structures of the CNOT2 and CNOT3 NOT-box domains. (**a**) Structure of the CNOT3 NOT-box dimer. Monomers are colored in dark and light green. Secondary-structure elements are labeled. (**b**) Structure of the CNOT3 NOT-box monomer. (**c**,**d**) Structure of the CNOT2 NOT-box dimer in two orientations. Monomers are colored purple and rose. The three α-helices involved in the domain swap are labeled in **d**. (**e**,**f**) Superposition of CNOT2 (purple) and CNOT3 (green) monomers. The similarity of the β-barrels and the open (CNOT2) and closed (CNOT3) conformations of the N-terminal α-helices are highlighted (**e**). A cross-connection of the two peptide backbones at the dashed line in **d** results in a compact model of the CNOT2 NOT box that is highly similar to the CNOT3 NOT box (**b**,**f**).

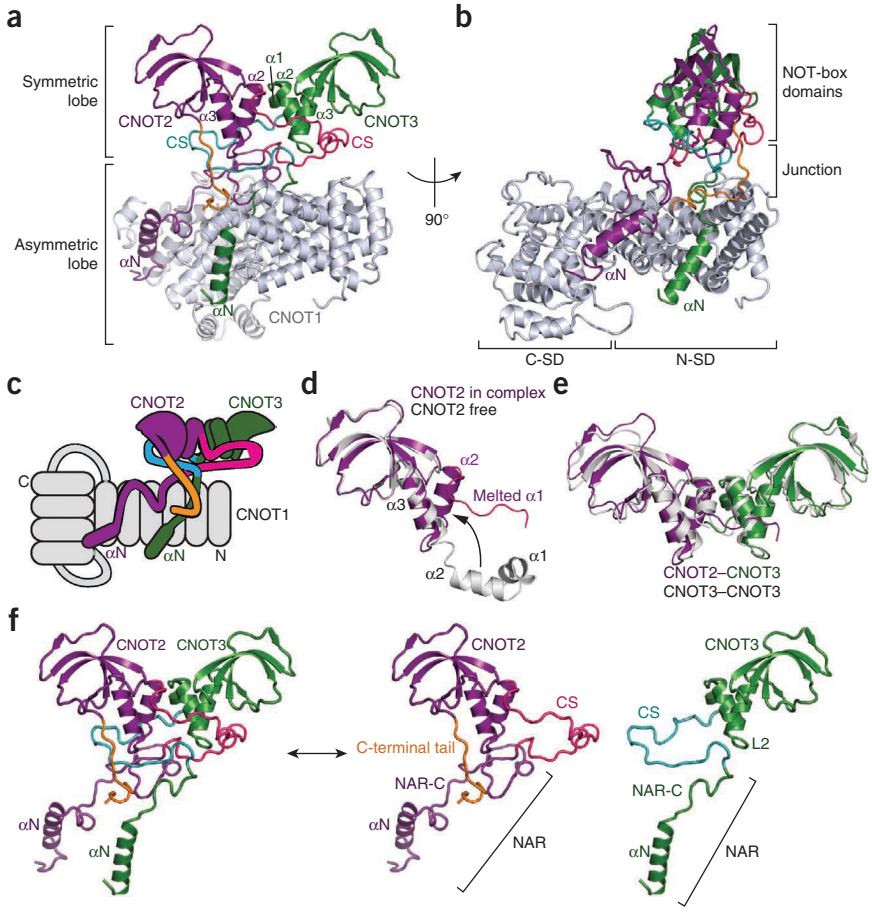


Figure 3 Structure of the CNOT1–CNOT2–CNOT3 ternary complex. (**a**,**b**) Structural overview. CNOT2 (purple) and CNOT3 (green) form a heterodimer through their NOT-box domains and preceding connector sequences (CS; magenta and cyan, respectively). The NOT1 anchor regions (NAR) tether the heterodimer onto the surface of CNOT1 (gray). They consist of an N-terminal α-helix (αN) and a NAR C-terminal region (NAR-C). The CNOT2 C-terminal tail is shown in orange. (**c**) Schematic drawing of the CNOT1–CNOT2–CNOT3 structure with colors as in **a**. (**d**) Superposition of the CNOT2 NOT-box domain in the complex (purple) with the isolated domain (gray). (**e**) Superposition of the CNOT2–CNOT3 heterodimer (purple and green) with the CNOT3 NOT-box homodimer (gray). (**f**) Structures of CNOT2 and CNOT3 from the ternary complex. The heterodimer (left) is taken apart (right) to show the extended conformation of the NAR and CS regions.

most of the contacts with additional contributions from helices α1 and loops L1. The contacts are mainly hydrophobic and lead to an interface of 560 Å^2.

The CNOT2 NOT-box domain forms a domain-swapped dimer

The CNOT2 NOT-box domain packs as a dimer of dimers (2.4-Å resolution, **Fig. 2c**,**d** and **Table 1**), consistent with its behavior in solution (**Supplementary Fig. 3d**,**e**). As expected on the basis of sequence homology (**Supplementary Fig. 3f**), the CNOT2 and CNOT3 NOT boxes adopt a similar fold (**Fig. 2c–f**). However, in the CNOT2 crystals the N-terminal helices adopt an extended (open) conformation (**Fig. 2c–e**), in which helices α1 and α2 of one monomer interact with helix α3 and the open β-barrel of the other monomer (**Fig. 2d**). Consequently, the dimers that result from this domain swap are structurally distinct from the CNOT3 homodimers, burying a large interface of 1,900 Å^2.

The ability of the NOT-box domains to homodimerize in two different ways and to adopt an open (CNOT2) or a closed (CNOT3) conformation is intriguing. With regard to the assembly of the NOT module, CNOT2 and CNOT3 may form heterodimers that correspond to one of the two crystallized arrangements. To discriminate between these two potential interaction modes, we determined the structure of the ternary complex, which also includes CNOT1.

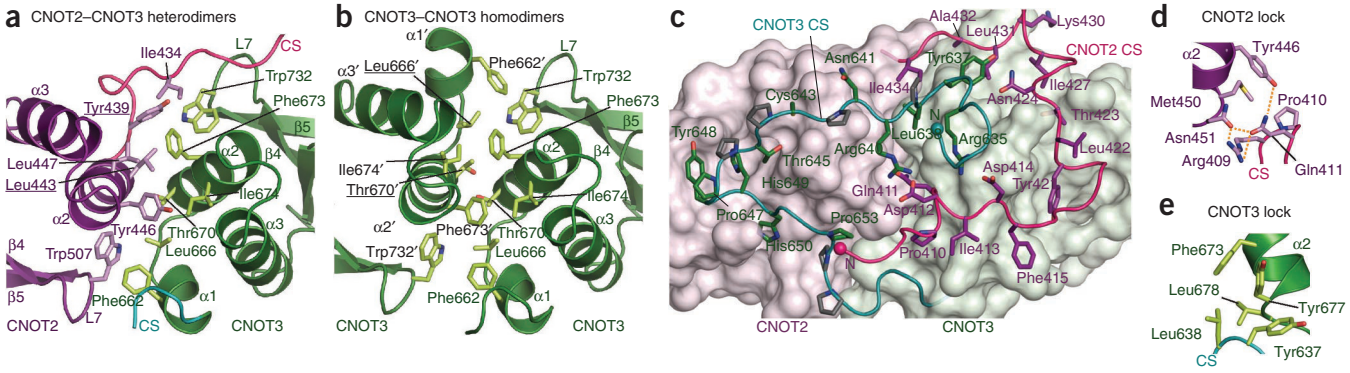


Figure 4 The CNOT2–CNOT3 heterodimerization interface. (**a**) Close-up of the interface between the NOT-box domains of CNOT2 and CNOT3, with interface residues shown as sticks. (**b**) The corresponding interface of the CNOT3 homodimer. Underlined residues indicate marked differences between the interfaces of the CNOT2–CNOT3 heterodimer and the CNOT3 homodimer. (**c**) Packing of the CNOT2 CSs (magenta) against the CNOT3 NOT-box domain (green surface) and of the CNOT3 CS (cyan) against the CNOT2 NOT-box domain (purple surface). Interacting residues are shown as sticks in purple (CNOT2) and green (CNOT3). Proline residues in CNOT3 CS are shown in gray. (**d**,**e**) Circular locks for CNOT2 (**d**) and CNOT3 (**e**). Residues locking back the N-terminus of the CS to its own NOT-box domain are drawn as sticks. For CNOT2, contacts are mediated by hydrogen bonds (orange). For CNOT3, contacts are hydrophobic.

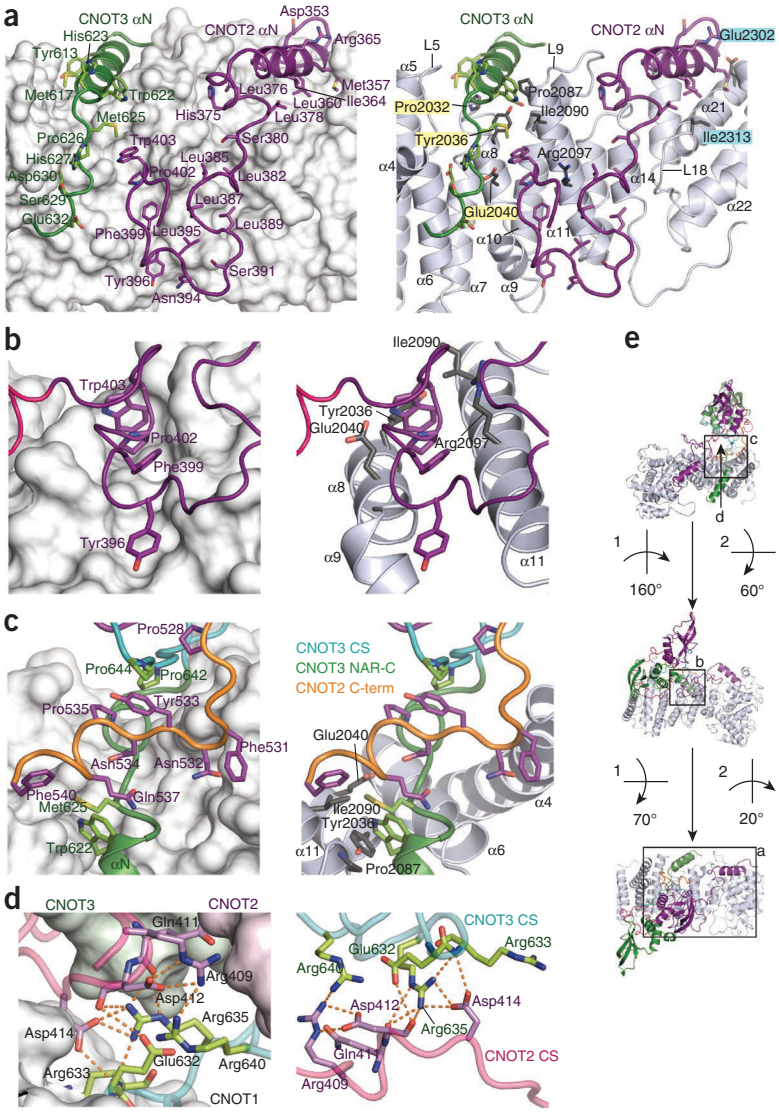


Figure 5 The interface between CNOT1 and the CNOT2–CNOT3 heterodimer. (**a**) Packing of the NARs (purple, CNOT2; green, CNOT3) against the surface of CNOT1 (gray, cartoon representation on the right). Interacting residues are shown as sticks and are labeled on the left for CNOT2 and CNOT3 and on the right for CNOT1. Highlighted labels correspond to CNOT1 mutants M1 (yellow) and M5 (cyan). (**b**) Close-up view of a hydrophobic canyon on the surface of CNOT1 filled with side chains of CNOT2. (**c**) Interactions of the CNOT2 C-terminal tail (C-term, orange) with CNOT3 and CNOT1. Labeling of interacting residues is as in **a**. (**d**) Close-up views of the junction between the symmetric and asymmetric lobes of the ternary NOT-module complex. (**e**) Overview of the trimeric complex for orientation. The rectangles indicate the views shown in **a**–**d**.

Crystal structure of the assembled NOT-module core

The core of the human NOT module comprising CNOT1 (residues 1565–2371), CNOT2 (residues 344–540) and CNOT3 (residues 607–753) was coexpressed in *E. coli* cells and co-purified. Limited proteolysis by thermolysin removed the N-terminal part of CNOT1, resulting in a stable fragment corresponding to the NOT1 SH domain (**Supplementary Fig. 4a**). Surprisingly, limited proteolysis did not affect the N-terminal extensions of the CNOT2-C and CNOT3-C fragments, although they are predicted to lack secondary structure. In the absence of CNOT1, thermolysin rapidly degraded CNOT2-C and CNOT3-C, a result suggesting that large portions of the two proteins are flexible and unfolded in isolation (**Supplementary Fig. 4a**,**b**).

The 3.2-Å-resolution structure of the heterotrimeric complex (**Fig. 3**, **Supplementary Fig. 4c**–**f** and **Table 1**) contains CNOT1 (1842–2353), CNOT2 (350–540) and CNOT3 (607–748) in a stoichiometry of 1:1:1. The complex is organized into two lobes: an asymmetric lobe containing CNOT1 is fixed by an unusual junction to a roughly two-fold-symmetric lobe containing the NOT boxes of CNOT2 and CNOT3 (**Fig. 3a**–**c**).

The structure of human CNOT1 in the complex is similar to the structure of the isolated *Ct* NOT1-C (r.m.s. deviation of 1.99 Å over 397 equivalent Cα positions; **Supplementary Fig. 4g**), thus indicating that CNOT1 does not undergo major structural rearrangements upon binding the CNOT2–CNOT3 heterodimer. Similarly, the CNOT3 NOT-box domain in the complex adopts an almost identical conformation as does its isolated counterpart (r.m.s. deviation of 0.55 Å over 90 equivalent Cα atoms; **Supplementary Fig. 4h**). In contrast to the isolated structure, the CNOT2 NOT-box domain adopts the closed conformation in the ternary complex, and its helix α1 is partially melted (**Fig. 3d** and **Supplementary Fig. 4i**).

The symmetric lobe

CNOT2 and CNOT3 interact with each other through their NOT-box domains in the same arrangement that was observed for the CNOT3 homodimer (**Fig. 3e**,**f**), thereby leading to a highly symmetric heterodimer. Heterodimerization is mediated by hydrophobic interactions between helices α1 and α2 and loop L7 of both proteins. Importantly, the interface of the CNOT2–CNOT3 heterodimer exhibits a higher degree of complementarity than does the interface in the CNOT3 homodimer (**Fig. 4a**,**b**). This is especially apparent for CNOT2 residues Y439 and L443, which form knobs that fit neatly into the holes on the CNOT3 surface, whereas the corresponding CNOT3 residues L666 and T670 fill these holes less efficiently.

The heterodimer interface is extended by N-terminal sequences that we named connector sequences (CSs) and that wrap like clamps around the NOT-box domains of their partners (**Figs. 3f** and **4c** and **Supplementary Fig. 5a**,**b**). The CSs describe a full circle starting from their C termini and wrap around the L2 loops of their respective partners (**Fig. 3f**), inserting into the cleft formed between the β-barrel and the N-terminal α-helices (**Fig. 4c**). In the vicinity of their N termini, the CSs lock themselves back through side chain interactions with helix α2 of their own NOT-box domain (**Fig. 4c**–**e**). The side chain interactions at the locks are different for the two proteins. In CNOT2, polar residues are involved for the most part (interaction of CS residues R409, P410 and Q411 with Y446, M450 and N451 in helix α2; **Fig. 4d**), whereas in CNOT3, the interactions are mainly hydrophobic (interaction of CS residues Y637 and L638 with residues F673, Y677 and L678 in helix α2; **Fig. 4e**). Most remarkably,

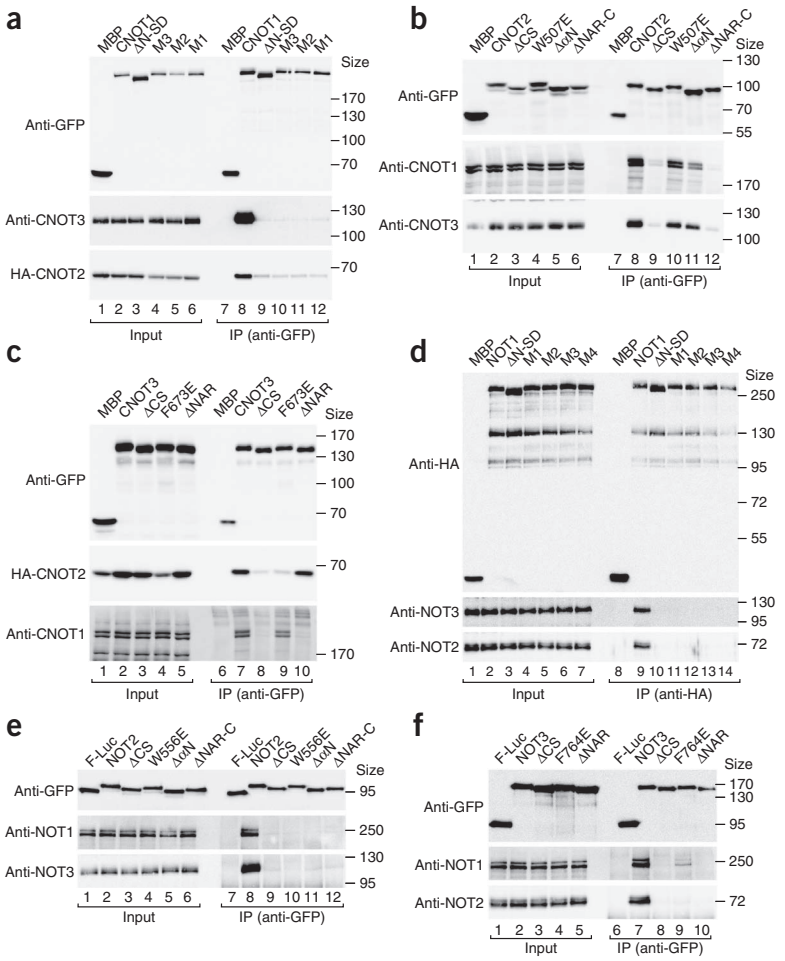


Figure 6 Mutagenesis of the NOT1-NOT2-NOT3 interfaces in human and *Dm* S2 cells. (**a**–**c**) Interaction of GFP-tagged CNOT1, CNOT2 or CNOT3 (either wild-type or the indicated mutants) with the indicated proteins (either endogenous CNOT3 and CNOT1 or CNOT2 tagged with human influenza hemagglutinin (HA)). GFP-tagged maltose binding protein (MBP) served as a negative control. (**d**) Interaction of HA-tagged *Dm* NOT1 (either wild-type or mutants) with the indicated endogenous proteins in S2 cells. HA-MBP served as a negative control. (**e**,**f**) Interaction of GFP-tagged *Dm* NOT2 or NOT3 (either wild-type or mutants) with the indicated endogenous proteins in S2 cells. GFP-tagged firefly luciferase (GFP-F-Luc) served as a negative control. In all panels, cell lysates were treated with RNase A before immunoprecipitation. IP, immunoprecipitated fraction. Size markers (kDa) are shown on the right of each panel. Original western blots shown in this figure can be found in **Supplementary Figure 8**.

the CSs are interlocked, thereby providing a topological constraint for the dissociation of the two proteins and suggesting a hierarchical assembly.

The CSs probably favor heterodimerization over homodimerization because they differ substantially in sequence and in length (**Figs. 3f** and **4c**). Notably, the CNOT3 CS is shorter and highly enriched in prolines (38%; **Fig. 4c** and **Supplementary Fig. 5a**,**b**). Thus, the interactions mediated by the CSs in the heterodimer are unlikely to occur in the context of the homodimers. This view is supported by the observation that the CNOT2 and CNOT3 fragments (containing the CSs in addition to the NOT boxes) aggregate and precipitate in isolation, thus indicating that the CSs do not participate favorably in the formation of homodimers. Upon coexpression, however, CNOT2-C and CNOT3-C remain soluble and form exclusively heterodimers (**Supplementary Fig. 3g**).

The asymmetric lobe

The CNOT2–CNOT3 heterodimer interacts predominantly with the conserved helical surface of the CNOT1 N-SD (**Fig. 3a**,**b**). The interaction does not use the NOT-box domains and involves only the NOT1 anchor regions (NARs) of both proteins, which consist of an N-terminal α-helix (αN) and a C-terminal region (NAR-C) entirely devoid of secondary structure (**Fig. 3f** and **Fig. 5a**–**d**). The αN helices insert into grooves on the lateral surface of CNOT1 (**Fig. 5a**). The extreme N-terminal residues of CNOT2 enter a cleft formed by loops L15 and L19 and helix α21 in the CNOT1 C-SD (**Fig. 5a**).

The NAR-Cs mold into the conserved helical surface of the CNOT1 N-SD (**Fig. 5a**). The short CNOT3 NAR-C crosses over the surface of helices α4, α6 and α8 of CNOT1, with which it interacts through hydrophobic and polar residues (**Fig. 5a**) before entering the junction. The long CNOT2 NAR-C zigzags across the CNOT1 surface and contains a series of aromatic residues (Y396, F399 and W403) that stick into the hydrophobic cleft between CNOT1 helices α8, α9 and α11 (**Fig. 5a**,**b**). Furthermore, residues 382–395 extensively interact with helix α11 of CNOT1 and exhibit several well-defined contacts (**Fig. 5a**).

The junction and the orientation of the two lobes

The NAR-Cs link to the CSs at the junction of the two lobes in the structure (**Fig. 3a**). The core of this junction is stabilized by hydrophilic interactions between the N-terminal ends of the CSs, which include residues from CNOT2 (R409, Q411, D412 and D414) and CNOT3 (E632, R633, R635 and R640; **Fig. 5d**). Furthermore, the relative orientation of the two lobes is stabilized by the C-terminal tail of CNOT2, which clamps the CNOT3 CS into a tunnel and hooks back onto CNOT1 (helix α11) while crossing over the CNOT3 NAR-C (**Fig. 5c**). Clearly, the interactions of the CNOT2 C-terminal tail must occur late in the assembly of the NOT module and impose additional topological constraints, thereby preventing the dissociation of CNOT3 in the presence of CNOT2. Surface conservation analysis indicated that the interfaces in the trimeric complex are conserved (**Supplementary Fig. 6a**–**e**), suggesting a similar assembly mode across multiple species.

In vivo assembly and mutagenesis of the NOT module

To determine the structural requirements for the assembly of the NOT module *in vivo*, we tested whether overexpressed protein variants are incorporated into endogenous CCR4–NOT complexes in human cells. To this end, we performed coimmunoprecipitation assays and used available anti-CNOT1 and anti-CNOT3 antibodies (**Fig. 6**). We included hemagglutinin (HA)-tagged CNOT2 in the transfection mixtures because of the lack of specific antibodies. Collectively, these experiments revealed the following observations.

First, GFP-CNOT1 interacted with HA-CNOT2 and endogenous CNOT3 (**Fig. 6a**, lane 8). As expected, a deletion of the CNOT1 N-SD (ΔN-SD) prevented these interactions (**Fig. 6a**, lane 9). Remarkably,

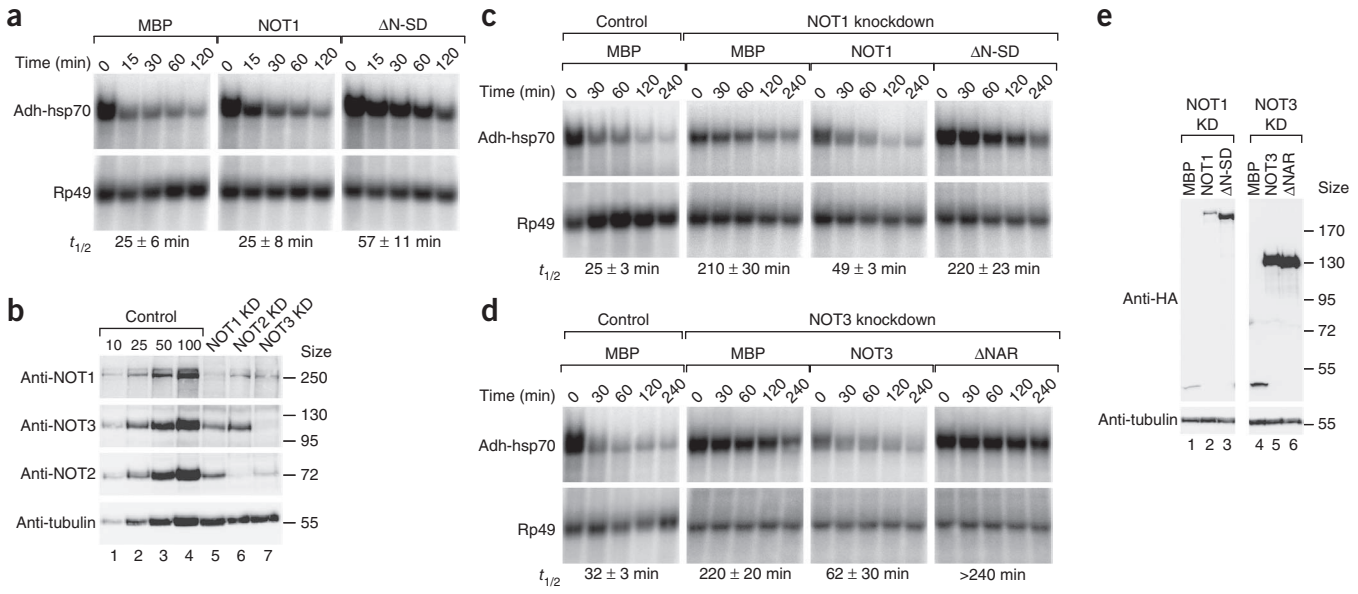


Figure 7 Effects on mRNA degradation. (**a**) Northern blot analysis showing the decay of the adh-hsp70 mRNA in control cells (expressing MBP) and in cells expressing NOT1 (either wild-type or mutant). The mRNA half-lives ($t_{1/2}$) ± s.d. calculated from the decay curves obtained from three independent experiments (**Supplementary Fig. 7d**) are indicated at bottom. Rp49 served as a loading control. (**b**) Western blot analysis of S2 cells depleted of NOT1, NOT2 or NOT3. Dilutions of control cell lysates were loaded in lanes 1–4 to estimate the efficacy of the depletion. Tubulin served as a loading control. KD, knockdown. Size markers (kDa) are indicated on the right. (**c**,**d**) Northern blots showing the decay of the adh-hsp70 mRNA in control cells (treated with GFP dsRNA and expressing MBP) or in cells depleted of NOT1 (**c**) or NOT3 (**d**) expressing MBP or the indicated proteins. The mRNA half-lives ($t_{1/2}$) ± s.d. calculated from the decay curves of three independent experiments (**Supplementary Fig. 7f,g**) are indicated at bottom. (**e**) Western blot showing the expression levels of the dsRNA-resistant, HA-tagged NOT1 and NOT3 proteins used in the complementation assays. Size markers (kDa) are indicated on the right. Original images of western and northern blots shown in this figure can be found in **Supplementary Figure 8**.

it was sufficient to substitute only two or three surface residues on CNOT1 that make direct contact with CNOT2 (mutants M5 and M6) or CNOT3 (mutant M1; **Fig. 5a** and **Supplementary Fig. 2**) to achieve a similar effect (**Fig. 6a**, **Supplementary Fig. 7a** and **Supplementary Table 1**). These CNOT1 mutants still interacted with CNOT7 (**Supplementary Fig. 7b**,**c**), a result suggesting that the mutations do not affect the CNOT1 fold. Consequently, subtle alterations of the CNOT1 binding surface are sufficient to prevent its interaction with CNOT2 and CNOT3 *in vivo*, a result consistent with a specific but rather weak, entropically costly interaction.

Second, GFP-CNOT2 interacted with endogenous CNOT1 and CNOT3 (**Fig. 6b**, lane 8). A deletion of the NAR-C region strongly reduced binding to endogenous CNOT1 (**Fig. 6b**, lane 12). This deletion also abolished binding to endogenous CNOT3 (**Fig. 6b**, lane 12), although a CNOT2 mutant lacking the entire NAR still interacted with CNOT3 *in vitro* (**Supplementary Fig. 1h**). This observation confirms that our assay detects the incorporation of CNOT2 into endogenous CCR4–NOT complexes, rather than detecting binary interactions alone. The results also suggest that there is no excess of free CNOT3 in the cell to interact with CNOT2. Conversely, a deletion of the CNOT2 CS also prevented the interaction with both CNOT1 and CNOT3 (**Fig. 6b**, lane 9). In contrast, a point mutation at the NOT-box interface (W507E) or the deletion of αN were ineffectual (**Fig. 6b**, lanes 10,11). These results highlight the importance of the CNOT2 CS and the NAR-C for NOT-module assembly.

Third, GFP-CNOT3 interacted with endogenous CNOT1 and with HA-CNOT2 (**Fig. 6c**, lane 7). As observed for CNOT2, the deletion of the CS prevents interaction not only with HA-CNOT2 but also with endogenous CNOT1 (**Fig. 6c**, lane 8), thus indicating that mutated GFP-CNOT3 is not incorporated into endogenous NOT modules. In contrast, a single amino acid substitution in CNOT3 helix α2 (F673E) disrupted HA-CNOT2 binding, but this mutant maintained interaction with endogenous CNOT1 (**Fig. 6c**, lane 9), a result suggesting that the CNOT3 NAR could be sufficient for incorporation into NOT modules even when the interaction with CNOT2 is disrupted. Consistent with this, the deletion of the NAR abrogated binding to endogenous CNOT1 (**Fig. 6c**, lane 10).

Finally, analogous experiments in *Dm* Schneider cells (S2 cells), demonstrated that all of the aforementioned mutations and deletions abolished the interaction of *Dm* NOT1, NOT2 and NOT3 with the other two endogenous partners (**Fig. 6d–f**). These results confirm the conclusion that endogenous NOT1, NOT2 and NOT3 are fully assembled into NOT modules and indicate that each subunit requires interaction with both of its partners to be incorporated into the NOT modules in *Dm* S2 cells. Together with the topological constraints seen in the structure, the results also suggest a highly coordinated and hierarchical assembly with built-in quality controls.

The integrity of the NOT module and mRNA degradation

To test whether a disturbed assembly of the NOT module has consequences on mRNA degradation, we used a heat-shock mRNA reporter containing the coding region of the *Dm* alcohol dehydrogenase gene (*Adh*) fused to the *hsp70* (official symbol *Hsp70Ab*) 3′ untranslated region (construct denoted adh-hsp70), which is sufficient to recapitulate *hsp70* mRNA decay[25] and results in a half-life of 25 ± 6 min (**Fig. 7a**).

In the first round of experiments, we overexpressed the NOT1 protein in S2 cells. We found that wild-type NOT1 had no effect on mRNA half-life (**Fig. 7a** and **Supplementary Fig. 7d**), whereas the expression of the NOT1 ΔN-SD mutant resulted in a two-fold increase of the mRNA half-life (57 ± 11 min; **Fig. 7a** and **Supplementary Fig. 7d**). Because this mutant still interacts with CAF1 (**Supplementary Fig. 7e**),

it probably sequesters the catalytic module into inactive complexes or it is assembled into defective complexes that are no longer recruited to the mRNA reporter.

In the second round of experiments, we depleted the individual subunits. Western blot analysis indicated that the depletion of single subunits of the ternary NOT1–NOT2–NOT3 complex co-depleted the other two subunits (**Fig. 7b**), in agreement with previous studies[11,13,15,25,26] (**Fig. 7b**). These observations demonstrate that the respective protein ratios are strictly controlled in the cell.

Because the depletions of NOT1 and NOT3 efficiently co-depleted the other two partners, we expected that such depletions would affect adh-hsp70 mRNA degradation in a similar manner. Consistent with this expectation and a previous study[25], the half-life of the hsp70 mRNA reporter in NOT1- and NOT3-depleted cells increased almost ten-fold relative to control cells (210 ± 30 and 220 ± 20 min, respectively; **Fig. 7c**,**d** and **Supplementary Fig. 7f**,**g**).

Next, we performed complementation assays, wherein NOT1 or NOT3 mutants were tested for their ability to restore mRNA degradation in cells that were depleted of the corresponding endogenous protein. Reintroduction of a double-stranded RNA (dsRNA)-resistant version of wild-type NOT1 partially restored the degradation of the reporter and reduced the mRNA half-life to 49 ± 3 min in NOT1-depleted cells (**Fig. 7c** and **Supplementary Fig. 7f**). In contrast, the NOT1 ΔN-SD mutant did not restore degradation, and the adh-hsp70 mRNA half-life (220 ± 23 min) was similar to that observed in NOT1-depleted cells (**Fig. 7c** and **Supplementary Fig. 7f**), even though expression was higher for the NOT1 mutant than for the wild-type (**Fig. 7e**).

Similarly, the reintroduction of a dsRNA-resistant version of wild-type NOT3 restored the degradation of the reporter and reduced the mRNA half-life to 62 ± 30 min in NOT3-depleted cells (**Fig. 7d** and **Supplementary Fig. 7g**). In contrast, the NOT3 ΔNAR mutant did not restore degradation (half-life > 240 min; **Fig. 7d**), even though the mutant was expressed at similar levels to wild-type NOT3 (**Fig. 7e**).

In summary, these experiments demonstrate a strict requirement for the cell to balance the levels of the NOT-module protein components. They also show that a properly assembled NOT module, as observed in the crystal structure, is essential for the activity and/or recruitment of the CCR4–NOT complex and hence is crucial for mRNA degradation.

DISCUSSION

Unusual assembly principles for the NOT module

A remarkable and unexpected structural feature of the NOT module is the role that the NAR and CS regions of CNOT2 and CNOT3 have in the assembly of the trimer. Particularly striking is that these regions not only hook up onto their prefolded molecular partners, as observed in assemblies such as the ribosome[27,28], but also stabilize each other and exclusively orchestrate the three-dimensional assembly of the entire NOT module. This mode of assembly is rather unusual and is clearly distinct from the classical principle of mutual recognition by complementary tertiary structures such as observed for the interaction of CAF1 with the NOT1 MIF4G domain[10,14]. In particular, the embracement of the NOT-box domains by the CSs, the formation of the junction and the adaptation of the NAR-Cs to the CNOT1 surface could not happen without disorder-to-order transition and cofolding[29].

According to the most general definition, the NAR and CS regions could be classified as intrinsically disordered, i.e., as protein sequences that are largely unstructured in the absence of their specific binding partners and that adopt their three-dimensional shape only upon binding[30,31]. This classification is based primarily on the lack of secondary-structure elements and of intramolecular contacts, which means that the relative orientation of the amino acids as observed in the complex will not be fixed in the absence of the binding partner. Consequently, the sequences are presumably disordered in isolation, as reflected by the susceptibility to proteolytic degradation (**Supplementary Fig. 4a**,**b**). However, sequence analysis does not identify the NAR and CS regions as typical intrinsically disordered regions[30,31].

In summary, we report the following new observations: first, even upon binding, the NAR and CS regions remain largely devoid of α-helices and β-strands, thus demonstrating that specific surface recognition and the formation of the structured core in the junction are possible without such elements. Second, rather than only threading along surface grooves, peptide regions of sufficient length, such as the CNOT2 NAR-C, can read out entire molecular surfaces by zig-zagging back and forth. Third, synergistic cofolding[29] of the CNOT2 and CNOT3 assembly peptides leads to a mutual stabilization and a specific hydrophilic core at the junction between the two lobes of the complex. Fourth, the interaction with these peptides is the sole determinant for NOT-module assembly and for the relative orientation of the two structured lobes, thus demonstrating that assembly peptides without secondary structure can guide and determine the shape of multimolecular complexes. Finally, topological constraints such as the described interlocks between the CSs increase the structural complexity and additionally stabilize the assembly.

Control of protein ratios and NOT-module assembly

The unexpected topological complexity of the NOT module suggests a requirement for a hierarchical and coordinated assembly of the individual subunits with the CNOT3 NAR binding the CNOT1 surface before CNOT2. Additional complexity arises from both CNOT2 and CNOT3 being able to form homodimers in the absence of their preferred binding partner, and these homodimers might compromise CCR4–NOT complex assembly or interfere with its function. Consequently, it is not surprising that the relative protein ratios of the NOT-module components are tightly controlled in the cell. In this context, the NAR and CS regions could have additional functions. They could interact with general chaperones or even specialized assembly factors that prevent nonspecific aggregation and/or promote the ordered incorporation into the CCR4–NOT complex. Alternatively, the NAR and CS regions could trigger an accelerated degradation of the proteins in the absence of their binding partners[32], owing to their susceptibility to cellular proteases, thus providing a simple mechanism to coordinate the expression of the subunits of the complex.

Recruitment of the CCR4–NOT complex to mRNA targets

The CCR4–NOT complex is assembled through the interaction of a catalytic module with the NOT module. Although the catalytic subunits catalyze deadenylation, the NOT module orchestrates the recruitment of these subunits to mRNA targets. This is achieved through interactions with RNA-associated proteins[2,17] (including GW182, Bicaudal-C, Nanos, CUP and Roquin), which have been shown to interact with NOT-module components[4,18–20,33–37]. The molecular details of these interactions remain unknown. In this context, our structural analysis reveals a highly conserved solvent-exposed surface on the NOT module that extends from the CNOT1 C-SD over the CNOT2 NAR to the CNOT3 NOT box (**Supplementary Fig. 6a**), and this surface probably provides a binding platform for conserved binding partners.

Remarkably, for the binding partners that have been characterized in more detail (GW182, Nanos and CUP)[4,18–20,35,36], the regions that

mediate binding to the NOT module are predicted to be unstructured. Thus their interactions with the NOT module may follow a similar structural principle as observed for the NAR and CS regions of CNOT2 and CNOT3. For example, the GW182 proteins interact with the CCR4–NOT complex through tryptophan-containing motifs[4,18–20] that probably bind in hydrophobic pockets along the CNOT1 scaffold and that may function similarly to the aromatic residues in the CNOT2 NAR-C. Understanding such interactions in molecular terms remains an important challenge for future studies.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Coordinates for the structures have been deposited in the Protein Data Bank under accession codes 4C0D (ternary complex of CNOT1–CNOT2–CNOT3), 4C0E (SH domain of *Ct* NOT1), 4C0F (NOT-box domain of CNOT2) and 4C0G (NOT-box domain of CNOT3).

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We are grateful to J. Su (Max Planck Institute for Developmental Biology) for providing the cDNA for the CNOT1 fragment (residues 1565–2371) and to E. Wahle (Martin Luther University Halle-Wittenberg) for the kind gift of anti–*Dm* NOT1–NOT3 antibodies. This work was supported by the Max Planck Society, by grants from the Deutsche Forschungsgemeinschaft (DFG, FOR855 and the Gottfried Wilhelm Leibniz Program awarded to E.I.) and the European Union Seventh Framework Program through a Marie Curie Fellowship to S.J. (FP7, 275343).

AUTHOR CONTRIBUTIONS

A.B., Y.C., T.R. and S.J. contributed equally to this work. A.B. and Y.C. purified, crystallized and solved the structures of CNOT3 and CNOT2 oligomers and of *Ct* NOT1. A.B. and Y.C. cloned, expressed and established the purification protocol for the ternary complex. T.R. purified and crystallized the ternary complex. T.R. and S.J. solved the structure of the ternary complex. A.B., Y.C., T.R., S.J. and O.W. collected and analyzed diffraction data. L.W. performed pulldowns and coimmunoprecipitations in human cells. D.K.-Ö. performed coimmunoprecipitations and functional assays in S2 cells. E.I. conceived of the project. E.I. and O.W. supervised the project. All authors contributed to the writing of the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Collart, M.A. & Panasenko, O.O. The Ccr4-Not complex. *Gene* **492**, 42–53 (2012).
2. Wahle, E. & Winkler, G.S. RNA decay machines: deadenylation by the Ccr4-Not and Pan2-Pan3 complexes. *Biochim. Biophys. Acta* **1829**, 561–570 (2013).
3. Bawankar, P., Loh, B., Wohlbold, L., Schmidt, S. & Izaurralde, E. NOT10 and C2orf29/NOT11 form a conserved module of the CCR4-NOT complex that docks onto the NOT1 N-terminal domain. *RNA Biol.* **10**, 228–244 (2013).
4. Chekulaeva, M. *et al.* miRNA repression involves GW182-mediated recruitment of CCR4–NOT through conserved W-containing motifs. *Nat. Struct. Mol. Biol.* **18**, 1218–1226 (2011).
5. Cooke, A., Prigge, A. & Wickens, M. Translational repression by deadenylases. *J. Biol. Chem.* **285**, 28506–28513 (2010).
6. Zekri, L., Kuzuoğlu-Öztürk, D. & Izaurralde, E. GW182 proteins cause PABP dissociation from silenced miRNA targets in the absence of deadenylation. *EMBO J.* **32**, 1052–1065 (2013).
7. Bai, Y. *et al.* The CCR4 and CAF1 proteins of the CCR4-NOT complex are physically and functionally separated from NOT2, NOT4, and NOT5. *Mol. Cell. Biol.* **19**, 6642–6651 (1999).
8. Lau, N.C. *et al.* Human Ccr4-Not complexes contain variable deadenylase subunits. *Biochem. J.* **422**, 443–453 (2009).
9. Liu, H.Y. *et al.* The NOT proteins are part of the CCR4 transcriptional complex and affect gene expression both positively and negatively. *EMBO J.* **17**, 1096–1106 (1998).
10. Basquin, J. *et al.* Architecture of the nuclease module of the yeast Ccr4-Not complex: the Not1-Caf1-Ccr4 interaction. *Mol. Cell* **48**, 207–218 (2012).
11. Ito, K., Takahashi, A., Morita, M., Suzuki, T. & Yamamoto, T. The role of the CNOT1 subunit of the CCR4–NOT complex in mRNA deadenylation and cell viability. *Protein Cell* **2**, 755–763 (2011).
12. Maillet, L., Tu, C., Hong, Y.K., Shuster, E.O. & Collart, M.A. The essential function of Not1 lies within the Ccr4-Not complex. *J. Mol. Biol.* **303**, 131–143 (2000).
13. Mauxion, F., Prève, B. & Séraphin, B. C2ORF29/CNOT11 and CNOT10 form a new module of the CCR4-NOT complex. *RNA Biol.* **10**, 267–276 (2013).
14. Petit, A.P. *et al.* The structural basis for the interaction between CAF1 nuclease and the NOT1 scaffold of the human CCR4-NOT deadenylase complex. *Nucleic Acids Res.* **40**, 11058–11072 (2012).
15. Russell, P., Benson, J.D. & Denis, C.L. Characterization of mutations in NOT2 indicates that it plays an important role in maintaining the integrity of the CCR4-NOT complex. *J. Mol. Biol.* **322**, 27–39 (2002).
16. Winkler, G.S., Mulder, K.W., Bardwell, V.G., Kalkhoven, E. & Timmers, H.T. Human Ccr4-Not complex is a ligand-dependent repressor of nuclear receptor-mediated transcription. *EMBO J.* **25**, 3089–3099 (2006).
17. Barckmann, B. & Simonelig, M. Control of maternal mRNA stability in germ cells and early embryos. *Biochim. Biophys. Acta* **1829**, 714–724 (2013).
18. Braun, J.E., Huntzinger, E., Fauser, M. & Izaurralde, E. GW182 proteins recruit cytoplasmic deadenylase complexes to miRNA targets. *Mol. Cell* **44**, 120–133 (2011).
19. Fabian, M.R. *et al.* miRNA-mediated deadenylation is orchestrated by GW182 through two conserved motifs that interact with CCR4–NOT. *Nat. Struct. Mol. Biol.* **18**, 1211–1217 (2011).
20. Huntzinger, E. *et al.* The interactions of GW182 proteins with PABP and deadenylases are required for both translational repression and degradation of miRNA targets. *Nucleic Acids Res.* **41**, 978–994 (2013).
21. Fabian, M.R. *et al.* Structural basis for the recruitment of the human CCR4–NOT deadenylase complex by tristetraprolin. *Nat. Struct. Mol. Biol.* **20**, 735–739 (2013).
22. Albert, T.K. *et al.* Isolation and characterization of human orthologs of yeast CCR4-NOT complex subunits. *Nucleic Acids Res.* **28**, 809–817 (2000).
23. Zwartjes, C.G., Jayne, S., van den Berg, D.L. & Timmers, H.T. Repression of promoter activity by CNOT2, a subunit of the transcription regulatory Ccr4-Not complex. *J. Biol. Chem.* **279**, 10848–10854 (2004).
24. Andrade, M.A., Petosa, C., O'Donoghue, S.I., Müller, C.W. & Bork, P. Comparison of ARM and HEAT protein repeats. *J. Mol. Biol.* **309**, 1–18 (2001).
25. Temme, C. *et al.* Subunits of the *Drosophila* CCR4-NOT complex and their roles in mRNA deadenylation. *RNA* **16**, 1356–1370 (2010).
26. Ito, K. *et al.* CNOT2 depletion disrupts and inhibits the CCR4–NOT deadenylase complex and induces apoptotic cell death. *Genes Cells* **16**, 368–379 (2011).
27. Brodersen, D.E. & Nissen, P. The social life of ribosomal proteins. *FEBS J.* **272**, 2098–2108 (2005).
28. Klein, D.J., Moore, P.B. & Steitz, T.A. The roles of ribosomal proteins in the structure assembly, and evolution of the large ribosomal subunit. *J. Mol. Biol.* **340**, 141–177 (2004).
29. Demarest, S.J. *et al.* Mutual synergistic folding in recruitment of CBP/p300 by p160 nuclear receptor coactivators. *Nature* **415**, 549–553 (2002).
30. Tompa, P. Intrinsically disordered proteins: a 10-year recap. *Trends Biochem. Sci.* **37**, 509–516 (2012).
31. Dyson, H.J. & Wright, P.E. Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* **6**, 197–208 (2005).
32. Prakash, S., Tian, L., Ratliff, K.S., Lehotzky, R.E. & Matouschek, A. An unstructured initiation site is required for efficient proteasome-mediated degradation. *Nat. Struct. Mol. Biol.* **11**, 830–837 (2004).
33. Temme, C., Zaessinger, S., Meyer, S., Simonelig, M. & Wahle, E. A complex containing the CCR4 and CAF1 proteins is involved in mRNA deadenylation in *Drosophila*. *EMBO J.* **23**, 2862–2871 (2004).
34. Chicoine, J. *et al.* Bicaudal-C recruits CCR4-NOT deadenylase to target mRNAs and regulates oogenesis, cytoskeletal organization, and its own expression. *Dev. Cell* **13**, 691–704 (2007).
35. Suzuki, A., Igarashi, K., Aisaki, K., Kanno, J. & Saga, Y. NANOS2 interacts with the CCR4-NOT deadenylation complex and leads to suppression of specific RNAs. *Proc. Natl. Acad. Sci. USA* **107**, 3594–3599 (2010).
36. Igreja, C. & Izaurralde, E. CUP promotes deadenylation and inhibits decapping of mRNA targets. *Genes Dev.* **25**, 1955–1967 (2011).
37. Leppek, K. *et al.* Roquin promotes constitutive mRNA decay via a conserved class of stem-loop recognition motifs. *Cell* **153**, 869–881 (2013).

ONLINE METHODS

Coimmunoprecipitation assays and western blotting. Plasmids expressing CCR4–NOT deadenylase subunits and coimmunoprecipitation assays in human and *Dm* S2 cells have been previously described[18]. HA- and GFP-tagged proteins were detected with horseradish peroxidase–conjugated anti-HA[18] (Roche 3F10; 1:3,000) and anti-GFP antibodies[18] (Roche 11814460001; 1:2,000), respectively. Endogenous human CNOT1 was detected with a polyclonal anti-CNOT1 antibody generated by immunization of rabbits with the purified untagged MIF4G domain of human NOT1 expressed in *E. coli*[14] (1:2,000). Endogenous CNOT3 was detected with a commercially available anti-CNOT3 antibody (Abcam ab55681; 1:1,000). The specificity of this antibody is shown in **Figure 6**. Endogenous *Dm* NOT1, NOT2 and NOT3 were detected with antibodies kindly provided by E. Wahle[25,33]. All western blots were developed with the ECL western blotting detection system (GE Healthcare).

***In vitro* pulldown assays.** The indicated GST- or MBP-tagged CNOT1, CNOT2 and CNOT3 fragments were expressed separately in BL21 (DE3) Star cells (Invitrogen) at 20 °C overnight. Cells were lysed in 10 mM HEPES, pH 7.6, 300 mM NaCl and 1 mM DTT supplemented with lysozyme (1 μg/ml), DNaseI (5 μg/ml) and protease inhibitors. Cell lysates were sonicated and cleared by centrifugation. The cleared supernatants containing the respective binding partners were mixed to obtain an approximately 1:1 ratio of the protein partners and incubated in binding buffer (10 mM HEPES, pH 7.6, 150 mM NaCl, 0.4% Triton X-100, 1 mM DTT, and protease inhibitors) for 20 min at 4 °C. After this, 50 μl (50% slurry) of Protino Glutathione agarose 4B beads (Macherey Nagel) or 50 μl (50% slurry) of amylose resin (New England BioLabs) was added to each sample, and incubation was continued for another hour at 4 °C with gentle rotation. Beads were washed three times with binding buffer. Bound proteins were eluted with 2× sample buffer and analyzed by SDS-PAGE.

His_6-tagged CNOT1 (residues 1565–2371) was expressed in *E. coli* BL21 (DE3) Star cells in LB medium at 20 °C overnight and purified through nickel-affinity chromatography followed by anion-exchange chromatography and finally size-exclusion chromatography, with a running buffer containing 50 mM HEPES, pH 8.0, 400 mM NaCl, 10% glycerol and 2 mM DTT. The purified protein (20 μg) was added to 200 μl of lysate expressing the respective MBP-tagged CNOT2 or CNOT3 and incubated in binding buffer (10 mM HEPES, pH 7.6, 300 mM NaCl, 1 mM DTT and protease inhibitors) for 20 min at 4 °C. After this, 50 μl (50% slurry) of amylose resin was added to each sample, and incubation was continued for another hour at 4 °C with gentle rotation. Beads were washed three times with binding buffer. Bound proteins were eluted with 2× sample buffer and analyzed by SDS-PAGE.

In the experiment shown in **Supplementary Figure 3g**, MBP-tagged CNOT2 and His_6-tagged CNOT3 were coexpressed in *E. coli*. The heterodimers were co-purified with MBP pulldown and subsequent Ni-affinity purification (lanes 1 and 2) or Ni-affinity purification and subsequent MBP pulldown (lanes 3 and 4). The stoichiometry of the complex was similar in all cases, thus indicating that the proteins heterodimerize in solution and that there was no excess of MBP-NOT2 or His_6-NOT3 homodimers.

Protein expression and purification. The DNA encoding *Ct* NOT1-C (residues 1676–2193) was amplified by PCR from a synthetic template (Gene Art, codon-optimized for *E. coli*) and inserted into *XhoI and BamHI* restriction sites of the pnYC vector, which provides an N-terminal His_6-tag followed by a TEV protease site[38]. His_6-tagged *Ct* NOT1-C was expressed in *E. coli* BL21 Star (DE3) cells at 20 °C overnight. After an initial Ni-affinity purification step (Hi-Trap chelating HP Nickel column; GE Healthcare), the protein was purified by gel filtration (HiLoad Superdex 200 26/60; GE Healthcare), in a buffer containing 10 mM HEPES, 200 mM NaCl and 2 mM DTT, and concentrated to 7 mg/ml.

The DNA encoding the CNOT2 NOT-box domain (residues 429–540) was amplified from total human cDNA and inserted into the *Nde*I and *BamH*I restriction sites of a pnYC vector downstream of a His_6-MBP tag. The fusion protein was expressed in *E. coli* BL21 Star cells at 37 °C for 5 h and purified over a Ni-NTA column. After cleavage by HRV3C protease at 4 °C overnight, the His_6-MBP tag was removed from the solution by binding to amylose resin and subsequent size-exclusion chromatography (HiLoad 16/60 Superdex 75; GE Healthcare). Finally, the protein was purified with a MonoQ GL 15/50 column, dialyzed overnight into 10 mM CHES, pH 9.0, 100 mM NaCl and 1 mM DTT and concentrated to 3.5 mg/ml.

The DNA encoding the CNOT3 NOT-box domain (residues 656–753) was amplified by PCR using total human cDNA as template and inserted into the *Nco*I and *KpnI* sites of the pETM60 vector, which provides an N-terminal His_6-NusA tag (Novagen). The fusion protein was expressed in *E. coli* BL21 Star cells at 20 °C overnight. After initial purification with Ni-NTA affinity chromatography, the His_6-NusA tag was cleaved by TEV protease at 4 °C overnight. CNOT3 was separated from the tag by size-exclusion chromatography (HiLoad 16/60 Superdex 75, GE Healthcare) and concentrated to 4.5 mg/ml in 10 mM CHES, pH 9.0, 150 mM NaCl and 1 mM DTT.

To obtain selenomethionine-substituted proteins, the CNOT2 NOT-box, CNOT3 NOT-box and *Ct* NOT1-C proteins were expressed in minimal medium supplemented with selenomethionine[39] and purified as described for the respective native proteins.

The trimeric complex consisting of CNOT1 (residues 1565–2371), CNOT2 (residues 344–540) and CNOT3 (residues 607–753) was coexpressed in *E. coli* BL21 (DE3) Star cells in LB medium at 20 °C overnight. CNOT1 was expressed with an N-terminal His_6-tag cleavable with TEV protease. CNOT2 and CNOT3 were expressed from a bicistronic plasmid, wherein CNOT2 contained an N-terminal MBP tag cleavable by HRV3C protease, and CNOT3 contained a noncleavable His_6 tag.

CNOT1, CNOT2 and CNOT3 were co-purified from crude lysates in lysis buffer (50 mM sodium phosphate, pH 7.5, 300 mM NaCl and 2 mM β-mercaptoethanol, supplemented with protease inhibitors, lysozyme and DNaseI) with amylose resin and eluted with lysis buffer supplemented with 25 mM maltose. The purified complex was digested overnight with HRV3C and TEV proteases during dialysis into imidazole buffer (10 mM Tris-HCl, pH 8.6, 200 mM NaCl, 2 mM β-mercaptoethanol, 10% glycerol and 20 mM imidazole). The complex was further purified by Ni-affinity chromatography and finally by size-exclusion chromatography (HiLoad Superdex 75 16/60 column; GE Healthcare) in gel-filtration buffer (10 mM Tris-HCl, pH 8.6, 200 mM NaCl and 2 mM DTT). The ternary complex was subjected to limited proteolysis for 4.5 h in proteolysis buffer (20 mM HEPES, pH 7.5, 100 mM NaCl, 50 μM $CaCl_2$ and 2 mM DTT) supplemented with 21 μg/ml thermolysin (Sigma-Aldrich) and separated from the protease and smaller CNOT1 fragments by gel filtration.

For the experiment shown in **Supplementary Figure 4b**, the ternary complex was assembled in 30 μl of proteolysis buffer with CNOT2 (residues 344–540) and CNOT3 (residues 607–753) at a concentration of 0.3 mg/ml and 0.63 mg/ml CNOT1 (residues 1833–2361). The binary complex of CNOT2 (residues 344–540) and CNOT3 (residues 607–753) was diluted in 30 μl of proteolysis buffer to a concentration of 0.3 mg/ml. After addition of 8 μl thermolysin (0.3 mg/ml) and incubation on ice, the reaction was stopped at different time points with 10 μl 5× SDS gel loading buffer and analyzed on a 15% SDS-PAGE gel.

Crystallization. Crystals of the native and selenomethionine-substituted *Ct* NOT1-C constructs were obtained by hanging-drop and sitting-drop vapor diffusion, respectively, at 18 °C and 22 °C. The best crystals for data collection and phase determination were obtained with selenomethionine-substituted protein, from a condition containing 0.1 M MES, pH 6.0, 0.18 M $MgCl_2$, 5% PEG 20000 and 10 mM proline with a 1:1 protein/reservoir ratio. Crystals were cryoprotected by addition of a final concentration of 25% glycerol to the reservoir solution and then flash-frozen in liquid nitrogen before data collection.

Crystals of the native CNOT2 NOT-box domain were grown by hanging-drop vapor diffusion at 18 °C. The protein solution was mixed in a 1:1 ratio with a reservoir solution containing 100 mM trisodium citrate, pH 5.6, 10% PEG 4000 and 10% isopropanol. Crystals were optimized by hair seeding. Before being flash frozen in liquid nitrogen, the crystals were cryoprotected in the reservoir solution supplemented with 15% or 20% glycerol (final concentration). Crystals of the selenomethionine-substituted CNOT2 construct were obtained by sitting-drop vapor diffusion at 22 °C over a reservoir of 100 mM HEPES, pH 7.5, 10% PEG 4000 and 10% isopropanol.

Crystals of the native and selenomethionine-substituted CNOT3 NOT-box domain were grown by hanging-drop vapor diffusion in a 1:1 protein/reservoir ratio, with the reservoir buffer containing 100 mM HEPES, pH 7.8, 13% PEG 4000, 2% glycerol and 12% isopropanol at 18 °C. Crystals were optimized by hair

doi:10.1038/nsmb.2681

seeding and were flash frozen in liquid nitrogen. Glycerol (supplemented to 20% in the reservoir solution) was used as a cryoprotectant.

Crystals of the CNOT1–CNOT2–CNOT3 complex were obtained once in a single condition. They grew after one day by sitting-drop vapor diffusion at 22 °C in a drop consisting of 0.2 µl of the protein complex solution (5 mg/ml) and 0.2 µl of the reservoir solution containing 0.1 M MES, pH 6.5 and 12% PEG 20000. The crystals were cryoprotected with the reservoir solution supplemented with glycerol (15% final concentration) and flash frozen in liquid nitrogen.

Data collection, structure solution and refinement. All diffraction data sets were recorded on a PILATUS 6M detector at the PXII beamline of the Swiss Light Source (SLS) at a temperature of 100 K. Data sets for different crystal forms of *Ct* NOT1-C were collected at a wavelength of 0.97925 Å with high redundancy and processed with XDS[40]. Selenium sites were identified with data sets from crystal form I (space group $P2_12_12_1$, 3.6 Å) with SHELX[41]. These sites were used to obtain phases, and an initial backbone model was built with PHENIX AutoSol[42]. This model was fed into Buccaneer[43] from the CCP4 package[44] for assigning and autobuilding side chains. The sequence register was confirmed by inspection of the anomalous difference Fourier maps from the selenomethionine data set, with peaks corresponding to the methionine positions. After several rounds of refinement in PHENIX[45] and manual model improvement in Coot[46], the model was used for molecular replacement (Phaser)[47] with a data set of crystal form II (space group $P12_11$, 3.2 Å). The structure was finalized by several cycles of manual building in Coot and refinement in PHENIX against this high-resolution data set. For *Ct* NOT1, 98.3% of all residues lie in the favored regions of the Ramachandran plot, with outliers totaling 0.2%.

Data for the native CNOT2 NOT-box domain were collected at a wavelength of 1.0 Å, whereas data from selenomethionine-labeled crystals were recorded at a wavelength of 0.97925 Å (peak). The best native crystal of CNOT2 led to a data set of 2.4-Å resolution with space group $P6_522$, whereas the best data set of selenomethionine-substituted protein reached a resolution of 3.4 Å with a space group of $P4_1$. Data sets were processed in XDS[40]. Initial selenium sites were identified by SAD with SHELX and refined with SHARP[48]. Additional sites were identified upon inspection of the anomalous difference density map and refined with SHARP. The resulting electron density map was used for initial model building with Buccaneer. The model was subsequently improved manually in Coot and refined with PHENIX. The high-resolution structure of the native protein was solved by molecular replacement (Phaser) with the structure of the selenomethionine-substituted protein as the search model. The structure of the native protein was finalized by iterative refinement cycles with PHENIX and model building in Coot.

The data sets for the CNOT3 NOT-box domain were collected at wavelengths of 0.9795 Å (peak) and 1.0000 Å (low-energy remote) and processed with XDS. Selenium sites were identified by SAD with SHELX. The heavy-atom model was further improved with SHARP as described for CNOT2. The model was initially built with Buccaneer and manually completed in Coot. Refinement was done in PHENIX against the low-energy remote data set, which had the highest resolution (2.4 Å). For the NOT-box domains of CNOT2 and CNOT3, there are 98.3% and 97.4% of all residues in the favored regions of the Ramachandran plot, respectively, and no outliers.

Diffraction data for the complex of CNOT1–CNOT2–CNOT3 were recorded at a wavelength of 0.99999 Å. The best data set was processed and scaled with XDS/XSCALE. Initial phase information was obtained by molecular replacement with Phaser, with the structure of *Ct* CNOT1-C as a search model. The initial electron density was improved by solvent flattening with PARROT[49] from the CCP4 package, and the previously determined models of the CNOT2 and CNOT3 NOT-box domains were manually placed into the density map with Coot. The CNOT1 model was subsequently rebuilt with the sequence of the human protein. Additional parts of the structure, which primarily consisted of the N-terminal extensions of the CNOT2 and CNOT3 NOT-box domains, were manually built in Coot. The model was subsequently improved by iterative cycles of refinement and building with PHENIX and Coot. For the ternary complex, 96.7% of all residues were in the favored regions of the Ramachandran plot, and there were no outliers. The correct stereochemical properties of all the structures were verified with MolProbity[50], and structure figures were generated in PyMOL (http://www.pymol.org/). The data collection and refinement statistics are summarized in **Table 1**.

Multiangle static laser light scattering (MALLS). The purified CNOT2 NOT-box and CNOT3 NOT-box proteins were diluted to a final concentration of 20 µM in their respective storage buffers. For both proteins, a volume of 100 µl was loaded onto an analytical gel-filtration column (Superdex 200 10/300, GE Healthcare) connected to the miniDAWN TREOS and Optilab rEX instruments (Wyatt Technologies). Samples were analyzed by static light scattering, and the absolute molecular weight of each protein was calculated from the MALLS data with ASTRA (Wyatt Technologies).

Functional assays in S2 cells. The adh-hsp70 reporter was described previously[51]. Cells were transfected with 300 ng of plasmid DNA for expression of the adh-hsp70 reporter per well in six-well plates. Knockdowns with dsRNA were performed as described previously[18]. For the measurement of the mRNA half-life, transfected cells were treated with actinomycin D (final concentration 5 µg/ml) 3 d after transfection and harvested at the indicated time points. Total RNA was isolated with TriFast (PeqLab Biotechnologies) and analyzed as described previously[51]. Adh-hsp70 mRNA levels were normalized to the levels of long-lived rp49 mRNA and were plotted against time.

Original images of gels and western and northern blots used in this study can be found in **Supplementary Figure 8**.

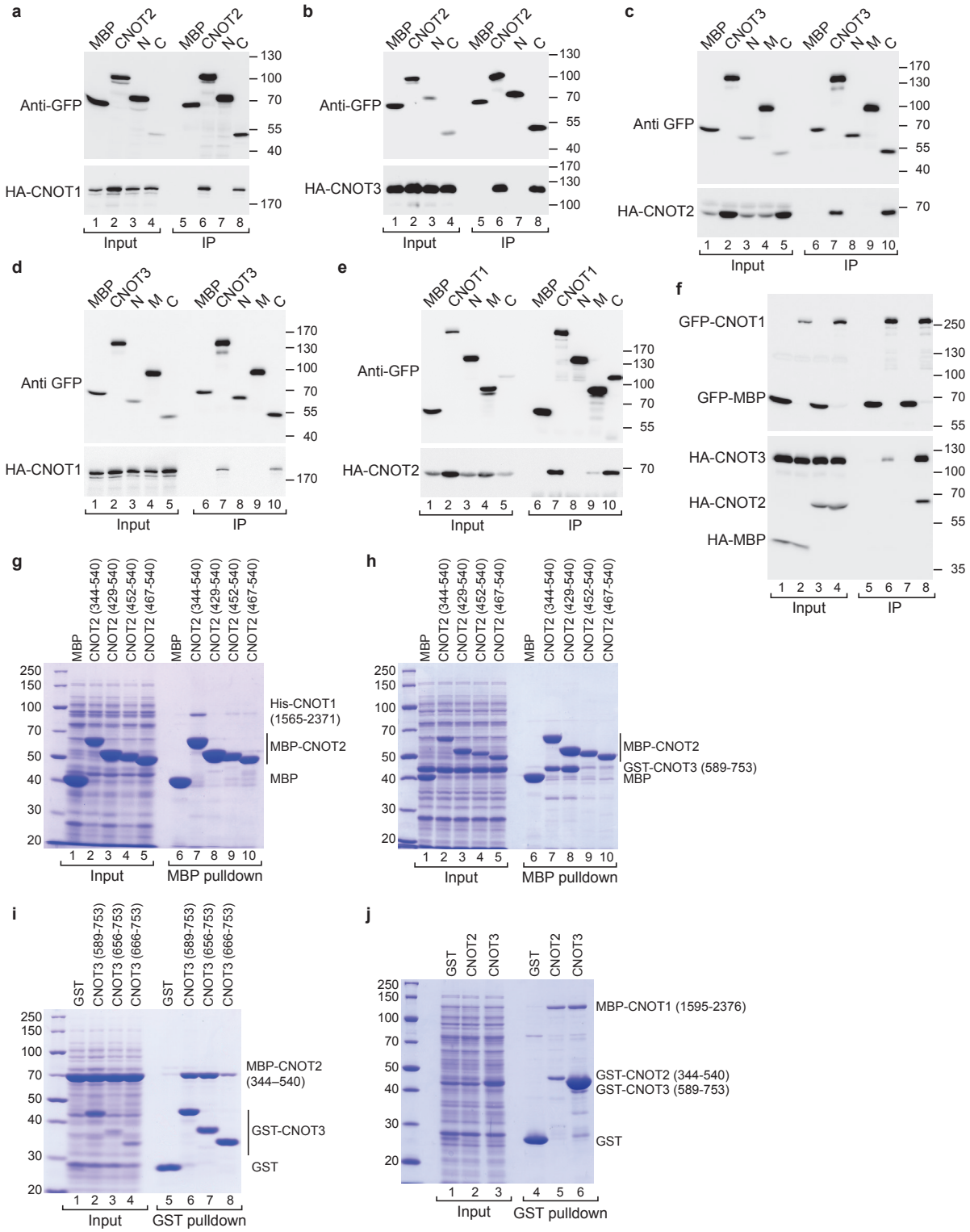
38. Diebold, M.-L., Fribourg, S., Koch, M., Metzger, T. & Romier, C. Deciphering correct strategies for multiprotein complex assembly by co-expression: application to complexes as large as the histone octamer. *J. Struct. Biol.* **175**, 178–188 (2011).
39. Doublié, S. Preparation of selenomethionyl proteins for phase determination. *Methods Enzymol.* **276**, 523–530 (1997).
40. Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
41. Sheldrick, G.M. Experimental phasing with SHELXC/D/E: combining chaintracing with density modification. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 479–485 (2010).
42. Terwilliger, T.C. *et al.* Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. *Acta Crystallogr. D Biol. Crystallogr.* **65**, 582–601 (2009).
43. Cowtan, K. The Buccaneer software for automated model building. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 1002–1011 (2006).
44. Winn, M.D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235–242 (2011).
45. Adams, P.D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
46. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
47. McCoy, A.J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
48. Vonrhein, C., Blanc, E., Roversi, P. & Bricogne, G. Automated structure solution with autoSHARP. *Methods Mol. Biol.* **364**, 215–230 (2007).
49. Zhang, K.Y.J., Cowtan, K. & Main, P. Combining constraints for electron-density modification. *Methods Enzymol.* **277**, 53–64 (1997).
50. Chen, V.B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21 (2010).
51. Behm-Ansmant, I., Gatfield, D., Rehwinkel, J., Hilgers, V. & Izaurralde, E. A conserved role for cytoplasmic poly(A)-binding protein 1 (PABPC1) in nonsense-mediated mRNA decay. *EMBO J.* **26**, 1591–1601 (2007).

Supplementary Information

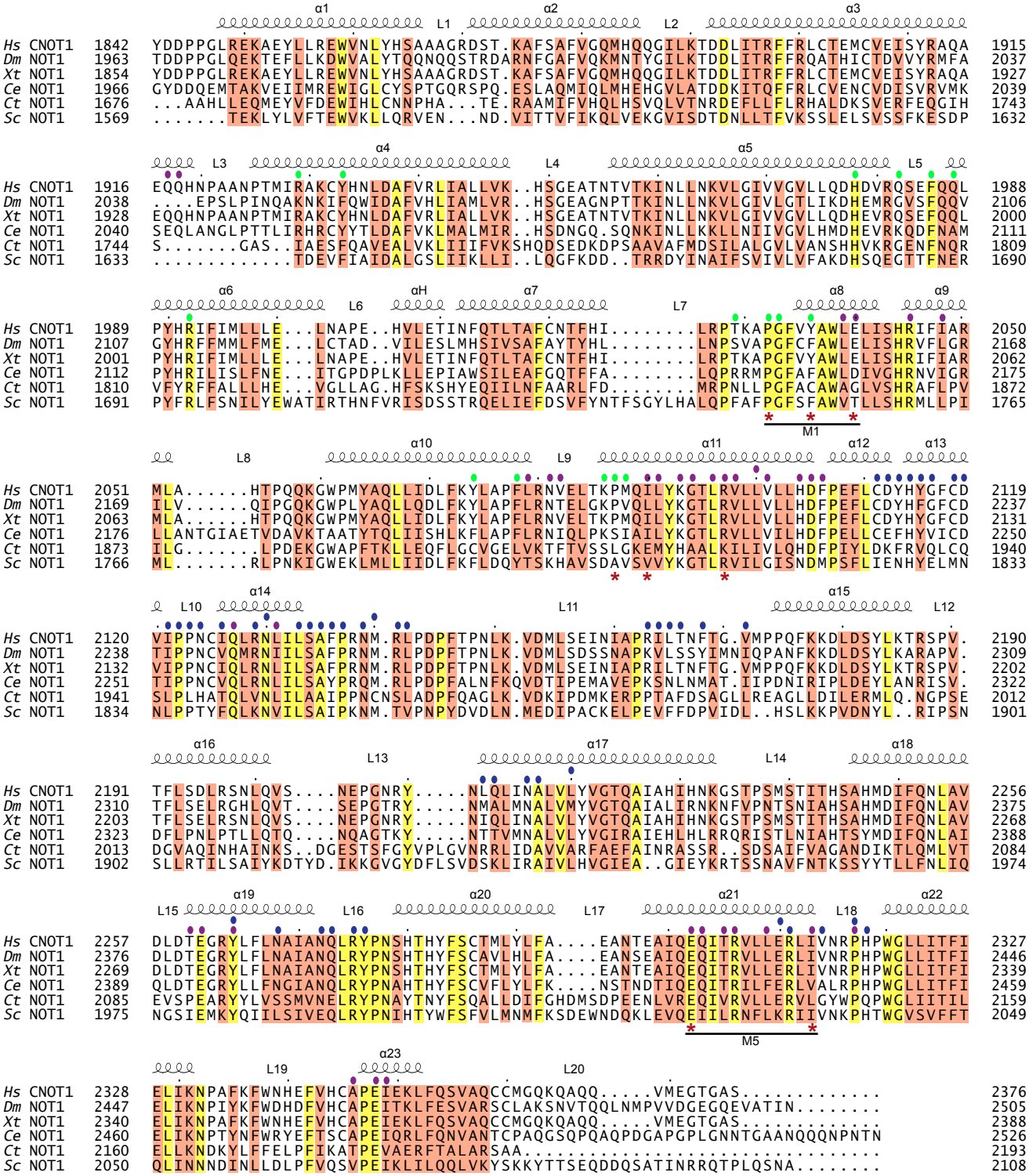
Structure and assembly of the NOT module of the human CCR4-NOT complex

Andreas Boland, Ying Chen, Tobias Raisch, Stefanie Jonas, Duygu Kuzuoğlu-Öztürk, Lara Wohlbold, Oliver Weichenrieder and Elisa Izaurralde

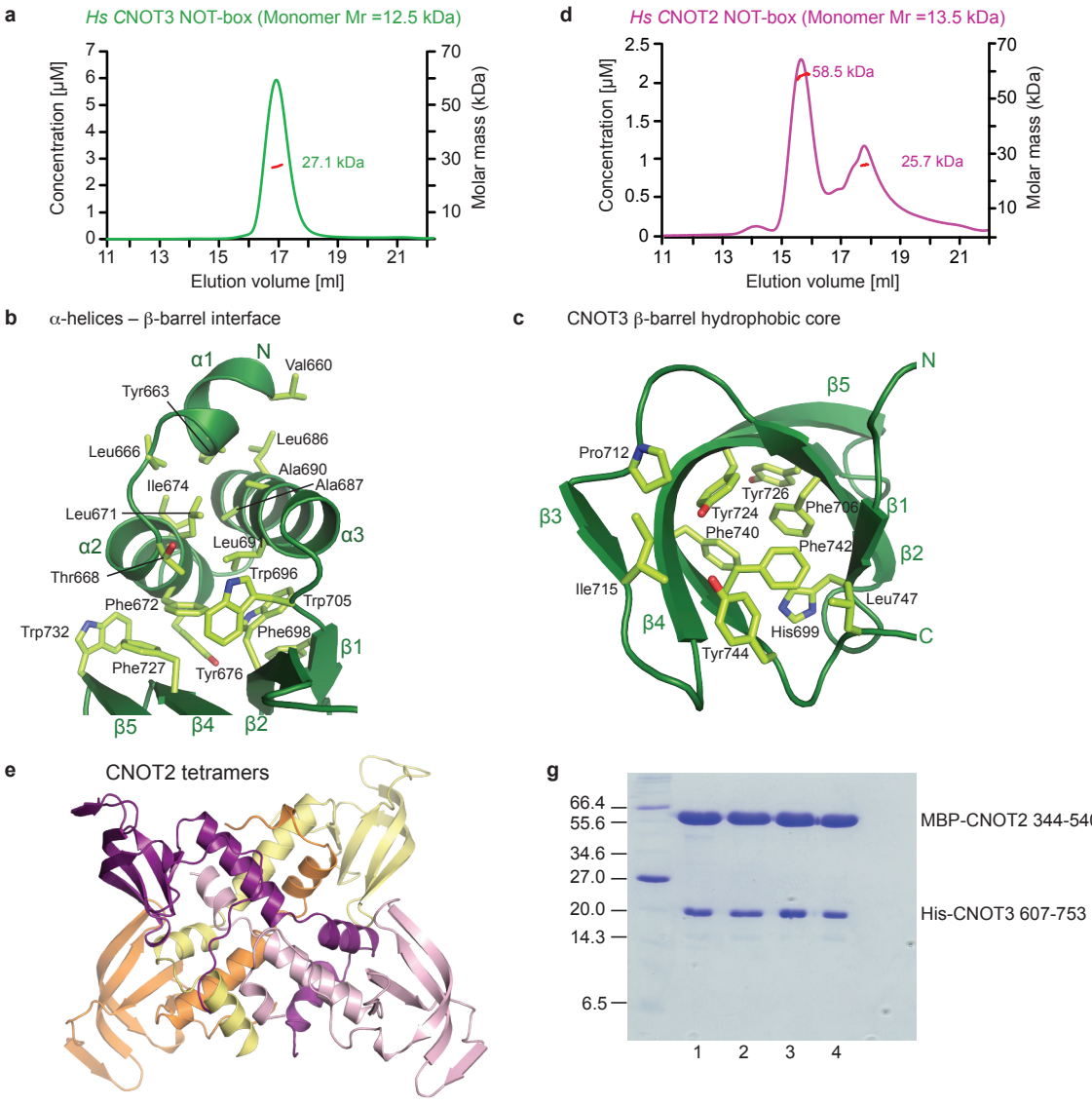
Nature Structural and Molecular Biology: doi:10.1038/nsmb.2681



Supplementary Figure 1 CNOT1, CNOT2 and CNOT3 interact via their C-termini. (**a**,**b**) Interaction between GFP-tagged CNOT2 (full-length or fragments) and HA-tagged CNOT1 (a) and CNOT3 (b). GFP-tagged MBP served as a negative control. In all panels, cell lysates were treated with RNase A. (**c**,**d**) Interaction between GFP-tagged CNOT3 (full-length or fragments) and HA-tagged CNOT2 (c) and CNOT1 (d). (**e**,**f**) Interaction between GFP-tagged CNOT1 (full-length or fragments) and HA-tagged CNOT2 (e) and CNOT3 (f). In panel (f), the interaction with CNOT3 was analyzed in the absence (lanes 2 and 6) or presence (lanes 4 and 8) of CNOT2. (**g**) Interaction between MBP-CNOT2 fragments and His_6-tagged CNOT1 (residues 1565–2371). (**h**) Interaction between MBP-CNOT2 fragments and GST-CNOT3 (residues 589–753). (**i**) Interaction between MBP-CNOT2 (residues 344–540) and GST-CNOT3 fragments. (**j**) Interaction between MBP-CNOT1 (residues 1595–2376) and GST-CNOT2 (residues 344–540) or GST-CNOT3 (residues 589–753).



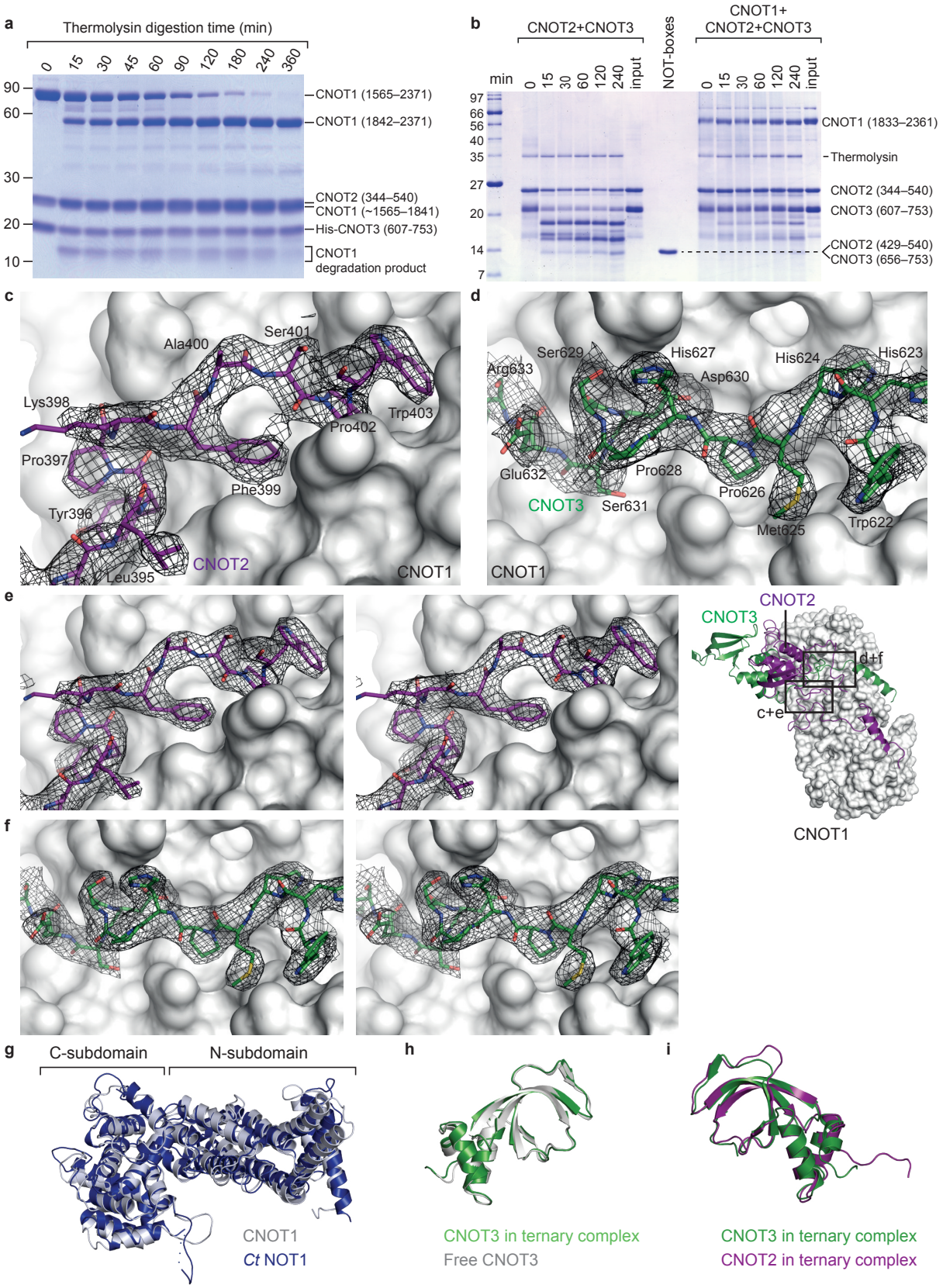
Supplementary Figure 2 Structure-based sequence alignment of the NOT1 superfamily homology (SH) domain. Secondary structure elements as determined from the Hs CNOT1 structure are shown above the alignment. Residues conserved in all aligned sequences are shown with a yellow background and residues with >70% similarity are highlighted in orange. Residues interacting with CNOT2 and CNOT3 are indicated by purple and green dots, respectively. Residues involved in CNOT1 subdomain interactions are indicated by blue dots. Residues mutated in this study are marked by red asterisks. Residues substituted in mutant M1 and M5 are indicated. The species abbreviations are as follows: *Hs* (*Homo sapiens*), *Dm* (*Drosophila melanogaster*), *Xt* (*Xenopus tropicalis*), *Ce* (*Caenorhabditis elegans*), *Ct* (*Chaetomium thermophilum*) and *Sc* (*Saccharomyces cerevisiae*).



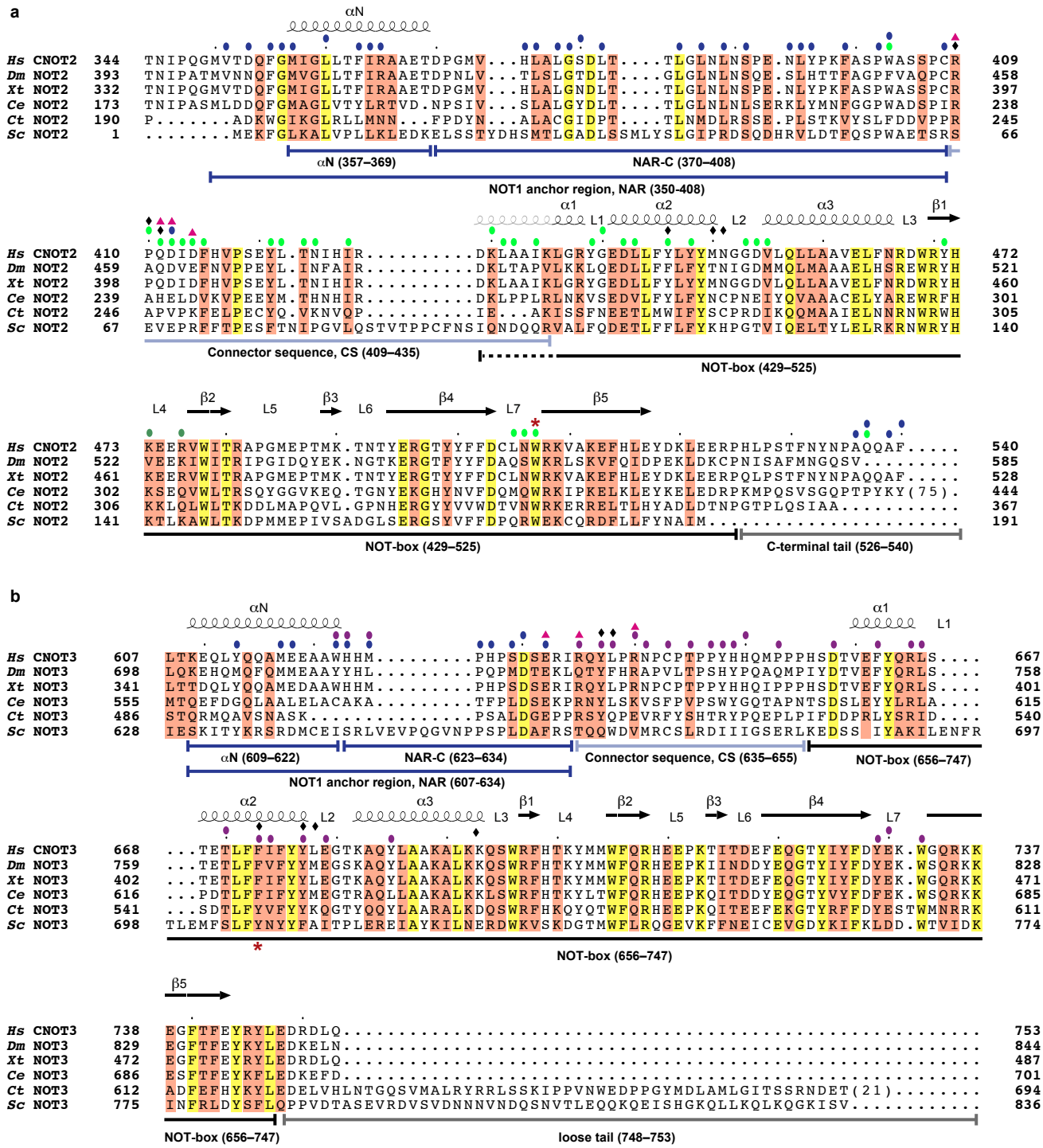
f

```
                    α1        α2          α3          β1   β2    β3   β4
                    L1   L2   L3   L4   L5   L6
Hs CNOT3    656 HSDTVEFYQRLSTETLFFIFYYLEGTKAQYLAAKALKKQSWRFHTKYMMWFQRH..EEPKTI.TDEFEQGTYIYFD 728
Dm NOT3     747 IYDTVEFYQRLSTETLFFVFYYMEGSKAQYLAAKALKKQSWRFHTKYMMWFQRH..EEPKII.NDDYEQGTYIYFD 819
Ct NOT3     529 IFDDPRLYSRIDSDTLFYVFYYKQGTYQQYLAARALKDQSWRFHKQYQTWFQRH..EEPKQI.TEEFEKGTYRFFD 601
Sc NOT3     693 YAKILENFRTLEMFSLFYNYYFAITPLEREIAYKILNERDWKVSKDGTMWFLRQ..GEVKFF.NEICEVGDYKIFK 765

Hs CNOT2    429 DKLAAIKLGRYGEDLLFYLYYMNGGDVLQLLAAVELFNRDWRYHKEERVWITRAPGMEPTMK.TNTYERGTYYFFD 503
Dm NOT2     478 DKLTAPVLKKLQEDLLFFLFYTNIGDMMQLMAAAELHSREWRYHVEEKIWITRIPGIDQYEK.NGTKERGTFYYFD 552
Ct NOT2     262 VQPIEAKISSFNEETLMWIFYSCPRDIKQQMAAIELNNRNWRWHKKLQLWLTKDDLMAPQVL.GPNHERGYYVVWD 336
Sc NOT2      97 IQNDQQRVALFQDETLFFLFYKHPGTVIQELTYLELRKRNWRYHKTLKAWLTKDPMMEPIVSADGLSERGSYVFFD 172

                    L7   β5
Hs CNOT3    729 YEK.WGQRKKEGFTFEYRYLEDRDLQ..............................................  753
Dm NOT3     820 YEK.WSQRKKEGFTFEYKYLEDKELN..............................................  844
Ct NOT3     602 LDD.WTVIDKINFRLDYSFLQPPVDTASEVRDVSVDNNN(33).............................  694
Sc NOT3     766 YESTWMNRRKADFEFHYKYLEDELVHLNTGQSVMALRYR(54).............................  836

Hs CNOT2    504 CLN.WRKVAKE.FHLEYDKLEERPHLPSTFNYNPAQQAF..................................  540
Dm NOT2     553 AQS.WKRLSKV.FQIDPEKLDKCPNISAFMNGQSV.....................................  585
Ct NOT2     337 TVN.WRKERRE.LTLHYADLDTNPGTPLQSIAA.......................................  367
Sc NOT2     173 PQR.WEKCQRD.FLLFYNAIM...................................................  191
```

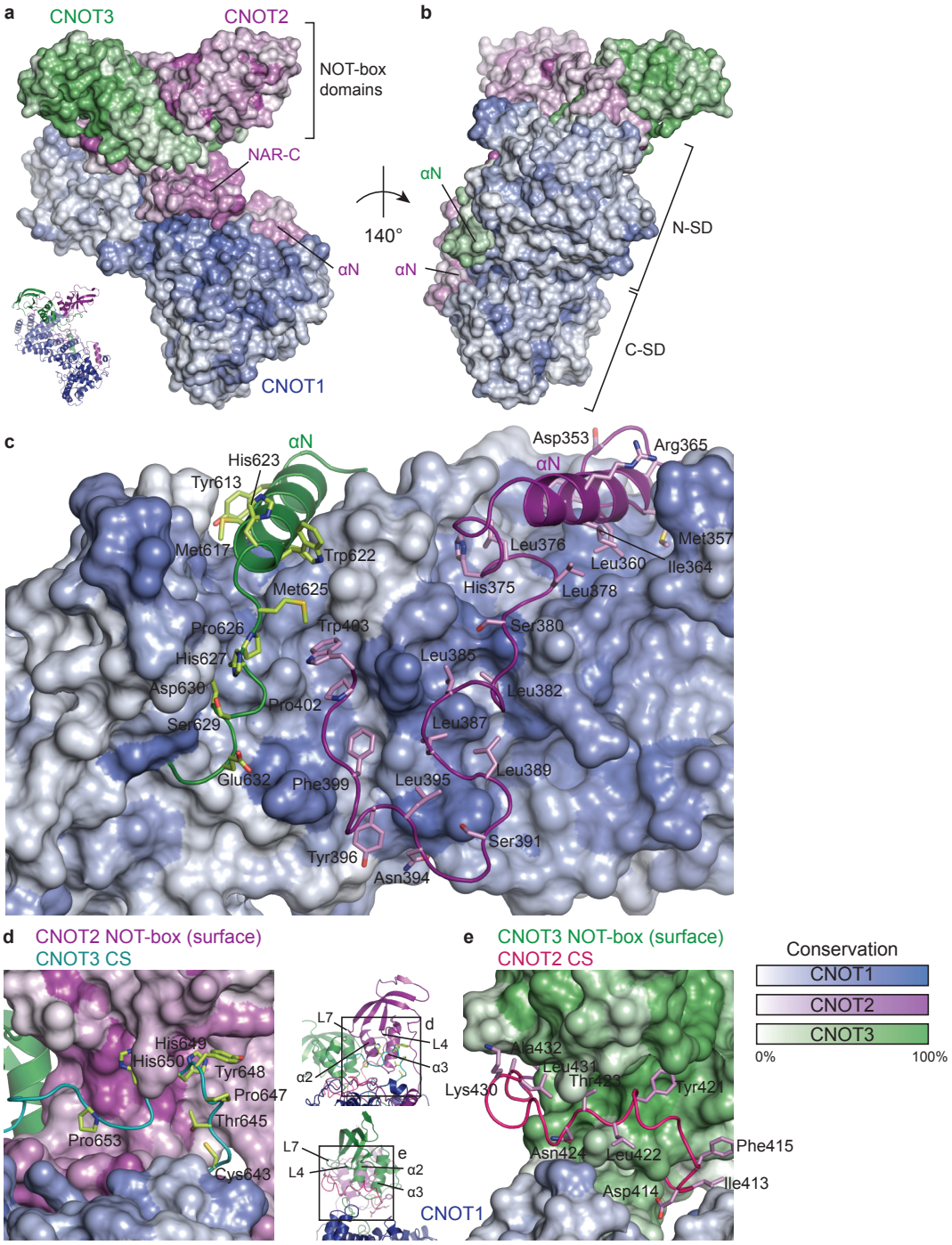
Supplementary Figure 3 CNOT2 and CNOT3 multimerize in solution. (**a**,**d**) MALLS analysis of the CNOT3 and CNOT2 NOT-box domains, respectively. The molecular weight of the protein in solution is indicated in the elution profile. (**b**) Interfaces between the CNOT3 NOT-box α-helices and the β-barrel. Residues along the interfaces are shown as sticks. (**c**) Hydrophobic core of the CNOT3 NOT-box β-barrel. (**e**) Structure of the CNOT2 tetramer as observed in the crystal. The tetramer consists of two pairs of dimers (orange-yellow vs. purple-rose) with a perpendicular orientation to each other. (**f**) Structure-based sequence alignments of the NOT2 and NOT3 NOT-box domains. Secondary structure elements as determined from the CNOT2 and CNOT3 structures are shown above the alignment. Residues conserved in all aligned sequences are shown with a yellow background, and residues with >70% similarity are highlighted in orange. A NOT2-specific insertion is boxed in purple. Black squares mark residues that form the interface between the NOT-Box N-terminal α-helices and the β-barrel. Gray squares mark residues that form the hydrophobic core of the β-barrel. The species abbreviations are the same as those in **Supplementary Fig. 2**. (**g**) Pulldowns showing that CNOT2-C and CNOT3-C exclusively form heterodimers in solution. MBP-tagged CNOT2 and His6-tagged CNOT3 were coexpressed in E. coli. The heterodimers were copurified using MBP pulldown followed by Ni-affinity purification (lanes 1 and 2) or Ni-affinity purification followed by MBP pulldown (lanes 3 and 4).



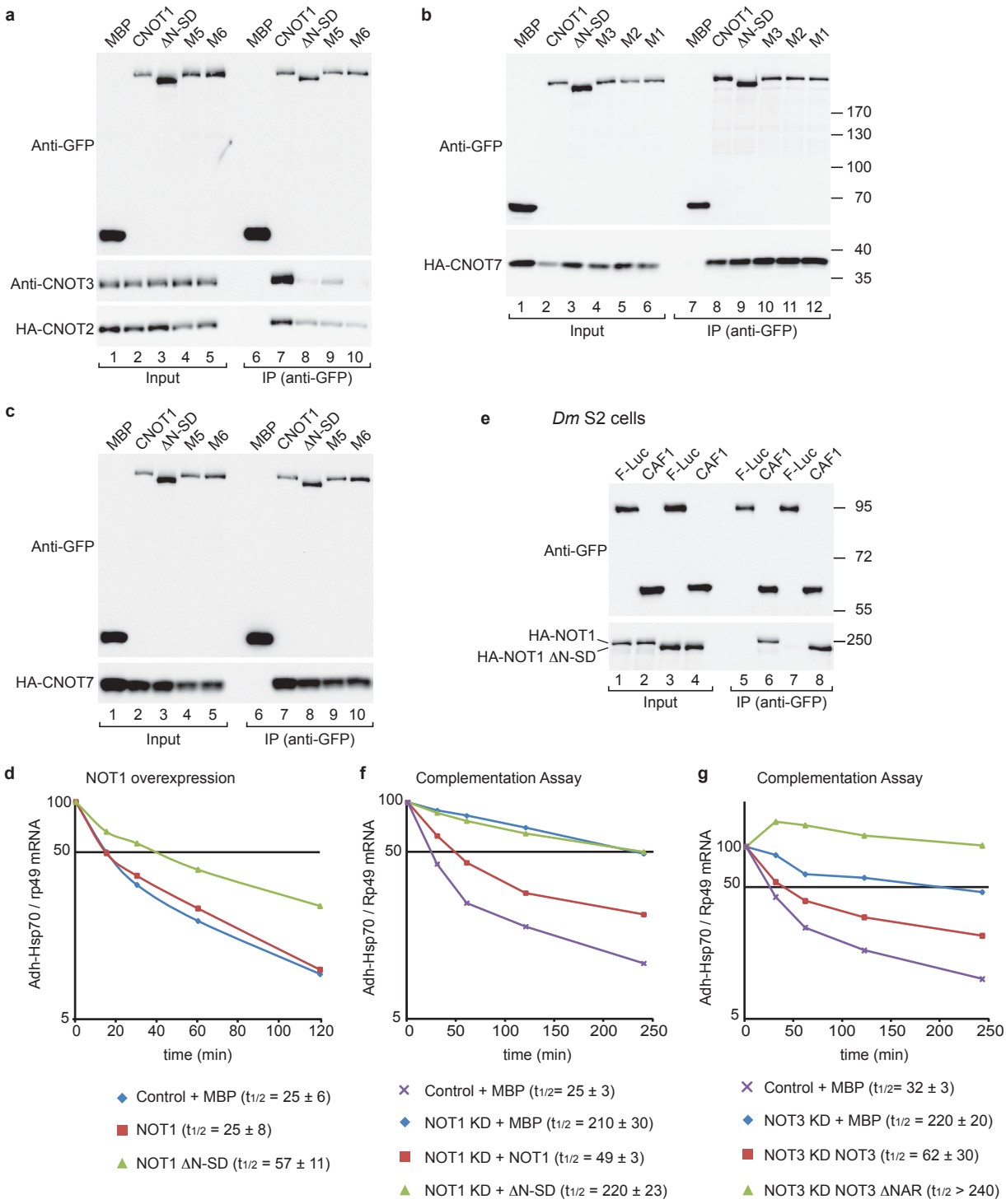
Supplementary Figure 4 Structure of the NOT module. (**a**) Time course of a limited proteolysis of the ternary complex between CNOT1 (residues 1565–2371), CNOT2 (residues 344–540) and CNOT3 (residues 607–753) by thermolysin. CNOT1 is digested into a stable C-terminal fragment corresponding to the SH and an N-terminal fragment (which comigrates with CNOT2), while CNOT2 and CNOT3 remain uncleaved. (**b**) Time course of a limited proteolysis of CNOT2 (residues 344–540) and CNOT3 (residues 607–753) by thermolysin in the absence or presence of CNOT1 SH (1833–2361). In the absence of CNOT1, CNOT2 and CNOT3 are digested into stable C-terminal fragments corresponding approximately to the NOT-boxes and the CSs, but they remain largely uncleaved in the presence of CNOT1 (**c**,**d**) Omit electron density map of the CNOT2 (c) and CNOT3 (d) NAR-Cs folded onto the CNOT1 surface. The electron density (black mesh, 2F0-FC of a composite omit map, calculated with Phenix.AutoBuild) is contoured at 1.0 σ. (**e**,**f**) Refined electron density of the CNOT2 (e) and CNOT3 (f) NAR-C regions in stereo view. The views correspond to panels (c) and (d), respectively. The electron density (black mesh, 2F0-FC map) is contoured at 1.0 σ. (**g**) Superposition of Ct NOT1 (blue) and Hs CNOT1 (gray) showing strong structure conservation. (**h**) Superposition of the CNOT3 NOT-box domain in isolation (gray) and in the complex (green). (**i**) Superposition of the NOT-box domains of CNOT2 (purple) and CNOT3 (green) as observed in the ternary complex.



Supplementary Figure 5 Structure-based sequence alignment of the CNOT2 and CNOT3. Secondary structure elements as determined from the CNOT2 and CNOT3 structure are shown above the alignment. Symbols are as described in **Supplementary Fig. 2**. Residues interacting with the respective partner proteins are indicated by dots above the alignments and are colored blue for CNOT1, purple for CNOT2 and green for CNOT3. Residues forming the lock are marked with black diamonds, and residues at the junction between the symmetric and asymmetric lobes of the ternary complex are marked with pink triangles. The species abbreviations are the same as those in **Supplementary Fig. 2**.



Supplementary Figure 6 Surface conservation of the trimeric complex. (**a**–**e**) The conservation scores of the individual residues are represented on the surface by color gradients from light (no conservation) to dark colors (100% conservation) for CNOT1 (blue), CNOT2 (purple) and CNOT3 (green). Conservation scores were calculated based on well-balanced multiple alignments covering all eukaryotic strata. (a) View from the CNOT1 surface that binds CNOT2 and CNOT3 (a) or the opposite surface (b). Conservation of the CNOT1 surface contacting CNOT2 and CNOT3 (c). The view is the same as that shown in **Fig. 5a**. (d) Conservation of CNOT2 surface residues contacting the CNOT3 connector sequence (CS). The CNOT3 residues involved in the interaction with CNOT2 are shown as sticks. (e) Conservation of CNOT3 surface contacting the CNOT2 connector sequence (CS). The CNOT2 residues involved in CNOT3 binding are shown as sticks.



Supplementary Figure 7 Mutagenesis of the NOT1-NOT2-NOT3 interfaces in human and Dm S2 cells. (**a**) Interaction of GFP-CNOT1 (either wild-type or the indicated mutants) with endogenous CNOT3 and HA-CNOT2. (**b**,**c**) Interaction of GFP-CNOT1 (either wild-type or the indicated mutants) with HA-CNOT7 in human cells. (**d**) The decay of the adh-hsp70 mRNA was monitored in control cells (expressing MBP) and in cells expressing NOT1 (either wild-type or mutant). Adh-hsp70 mRNA levels were normalized to the levels of long-lived rp49 mRNA and plotted against time. A representative Northern blot is shown in Fig. 7a. The mRNA half-lives (t1/2) ± standard deviations calculated from the decay curves obtained from three independent experiments are indicated. (**e**) Interaction of GFP-tagged Dm CAF1 with wild-type NOT1 or NOT1ΔN-SD in S2 cells. F-Luc-GFP served as a negative control. (**f**,**g**) The decay of adh-hsp70 mRNA was analyzed in control cells (treated with GFP dsRNA and expressing MBP) or in cells depleted of NOT1 or NOT3 and expressing MBP or the indicated proteins. Northern blots corresponding to the decay curves are shown in **Fig. 7c** and **7d**, respectively. Adh-hsp70 mRNA levels were normalized to the levels of rp49 mRNA and plotted against time. The mRNA half-lives (t1/2) ± standard deviations obtained from three independent experiments are indicated.

Fig 6a

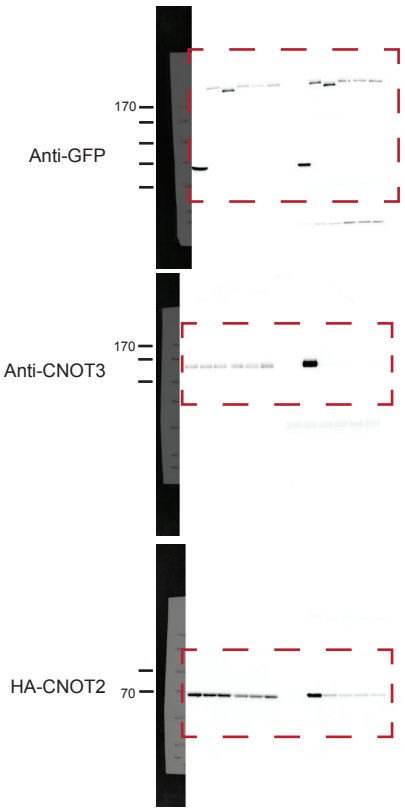


Fig 6b

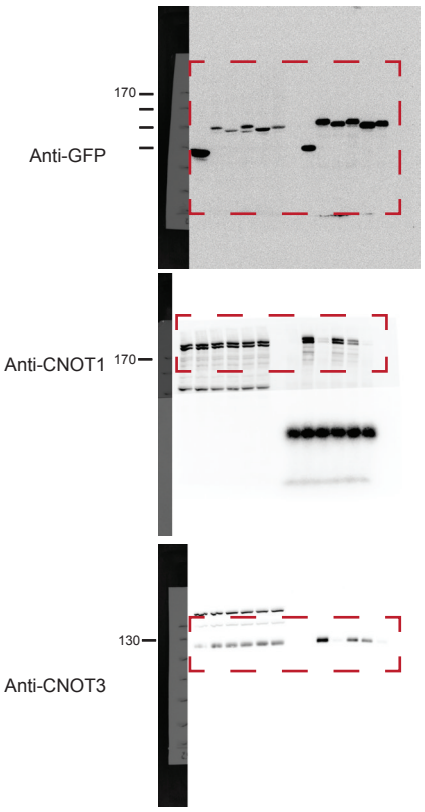


Fig 6c

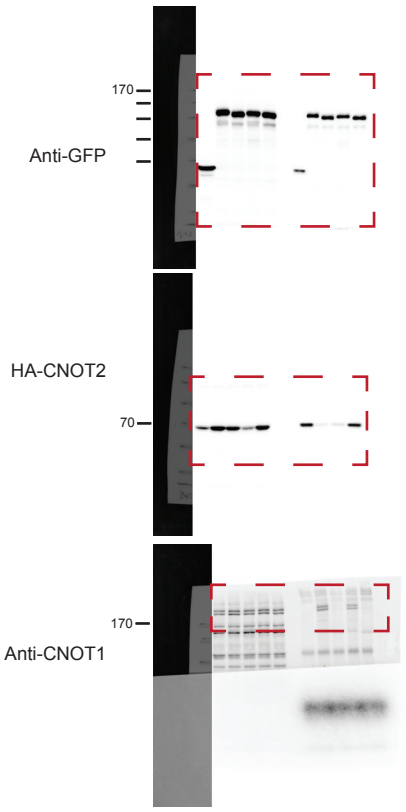


Fig 6d

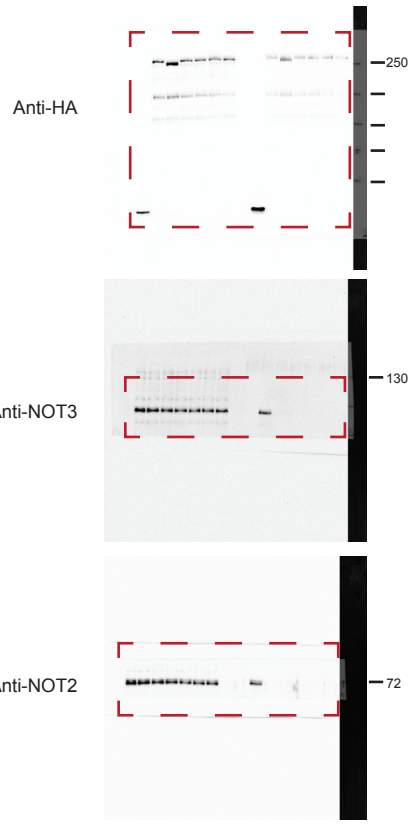


Fig 6e

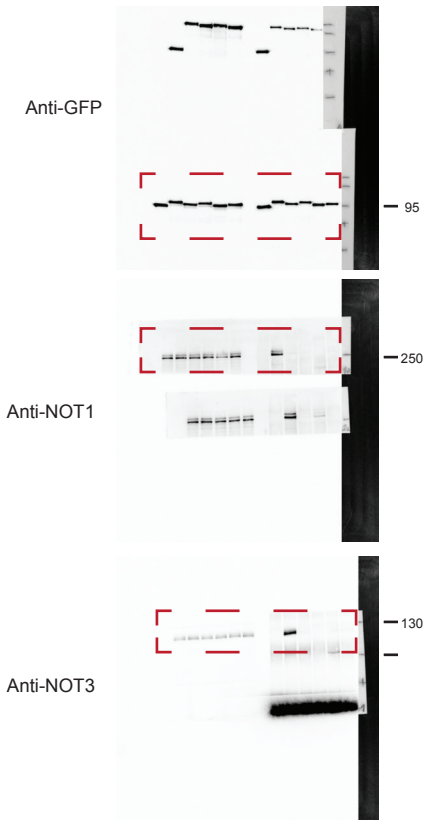
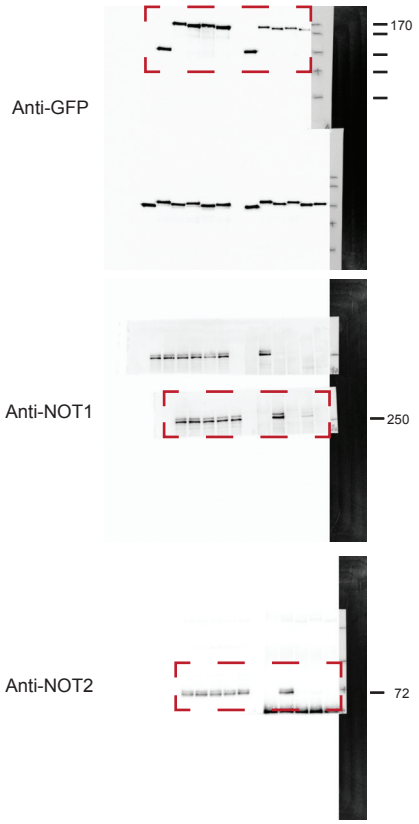


Fig 6f



Nature Structural and Molecular Biology: doi:10.1038/nsmb.2681

Fig 7a

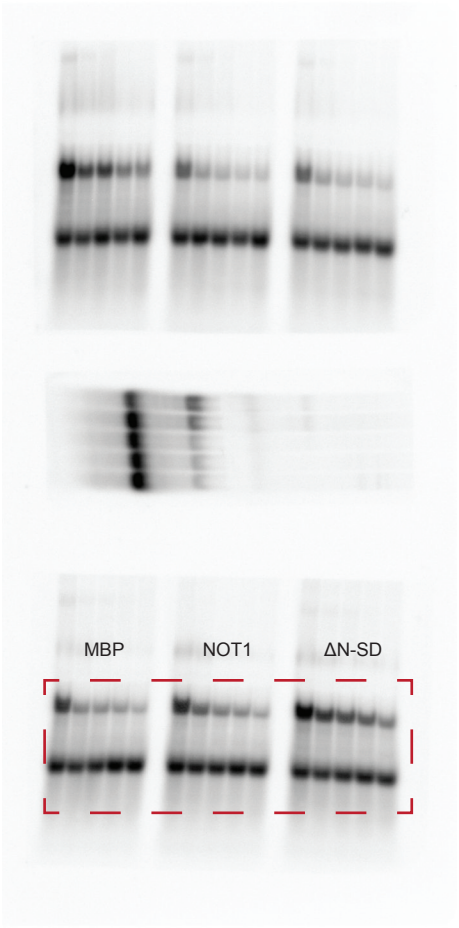


Fig 7d

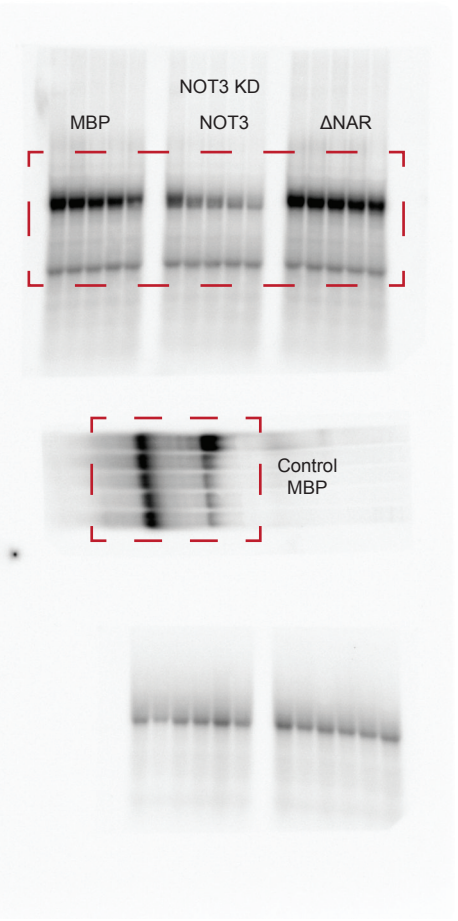
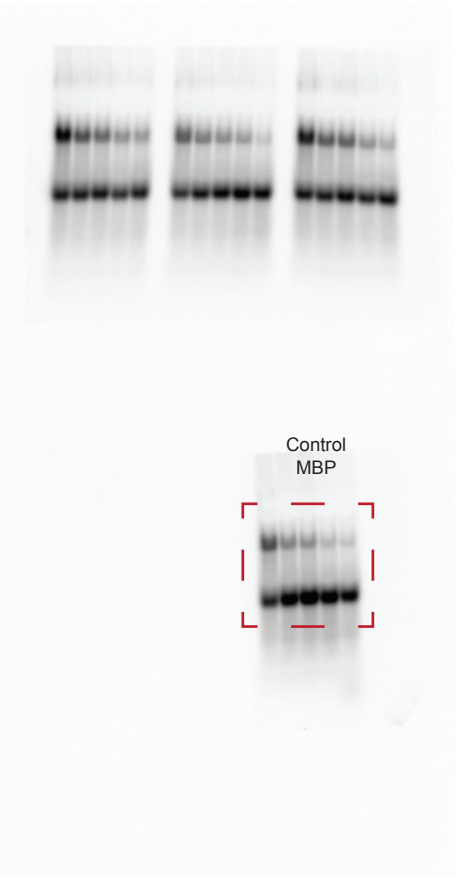


Fig 7c



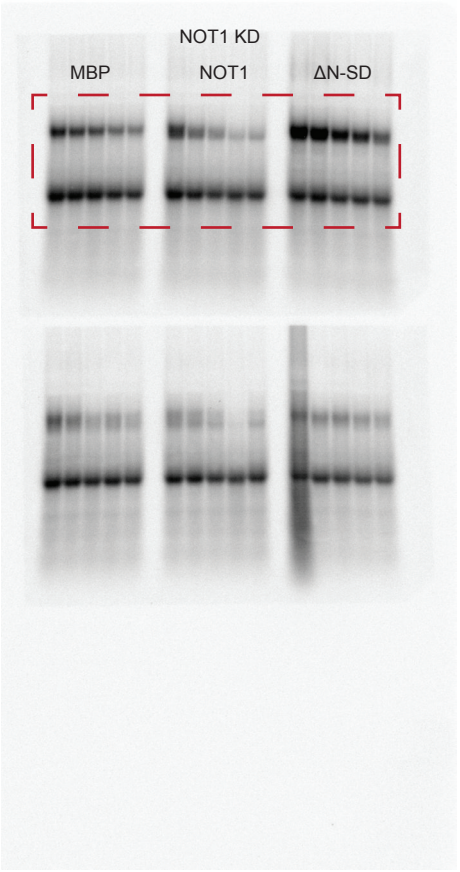


Fig 7b

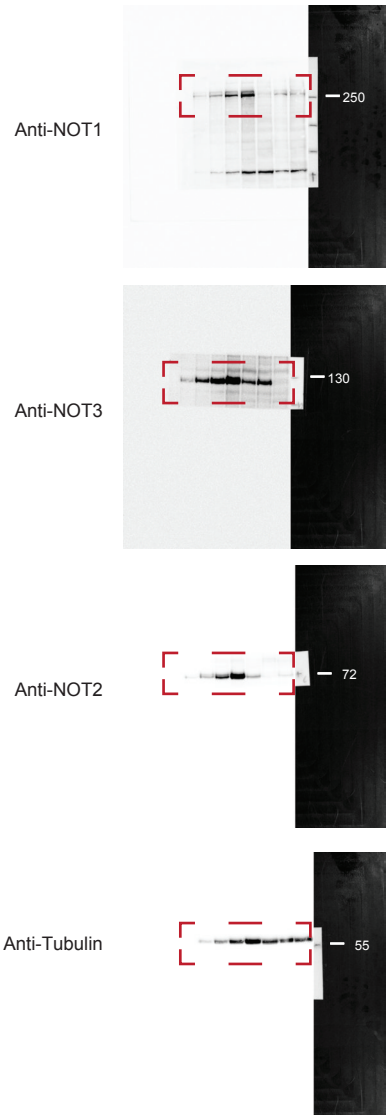
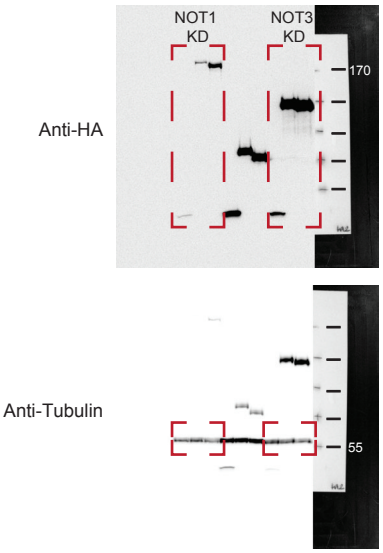


Fig 7e



Supplementary Figure 8 Original images of gels, western and northern blots used in this study.

Supplementary Table 1. Mutants used in this study

Name	*Dm* NOT1	*Hs* CNOT1
ΔN-SD	Δ1967-2251	Δ1844-2133
M1	P2150Y, F2152C, E2158A, P2205Q, L2208E	P2032Y, Y2036C, E2040A
M2	P2150Y, F2152C, E2158A, P2205Q, L2208E, E2421A	P2032Y, Y2036C, E2040A, P2087Q, I2090E, E2302A
M3	P2150Y, F2152C, E2158A, P2205Q, E2421A, L2208E, R2215D	P2032Y, Y2036C, E2040A, P2087Q, I2090E, R2097D, E2302A, I2313R
M4	P2150Y, F2152C, E2158A, P2205Q, E2421A, L2208E, R2215D, I2432R	
M5		E2302A+I2313R
M6		R2097D+I2313R
CNOT1-N		1–1089
CNOT1-M		1085–1605
CNOT1-C		1595–2376

Name	*Dm* NOT2	*Hs* CNOT2
ΔN-CS	Δ 458–482	Δ 409–432
ΔαN	Δ 393–429	Δ 344–380
ΔN-NAR-C	Δ 429–453	Δ 380–404
L7	W556E	W507E
CNOT2-N		1–352
CNOT2-C		344–540

Name	*Dm* NOT3	*Hs* CNOT3
ΔN-CS	Δ 723–745	Δ 632–654
ΔNAR	Δ 698–725	Δ 607–634
α2	F764E	F673E
CNOT3-N		1–241
CNOT3-M		239–590
CNOT3-C		589–753

Molecular Cell
Article

A DDX6-CNOT1 Complex and W-Binding Pockets in CNOT9 Reveal Direct Links between miRNA Target Recognition and Silencing

Ying Chen,[1,2] Andreas Boland,[1,2] Duygu Kuzuoğlu-Öztürk,[1,2] Praveen Bawankar,[1] Belinda Loh,[1] Chung-Te Chang,[1] Oliver Weichenrieder,[1,*] and Elisa Izaurralde[1,*]
[1]Department of Biochemistry, Max Planck Institute for Developmental Biology, Spemannstrasse 35, 72076 Tübingen, Germany
[2]Co-first authors
*Correspondence: oliver.weichenrieder@tuebingen.mpg.de (O.W.), elisa.izaurralde@tuebingen.mpg.de (E.I.)
http://dx.doi.org/10.1016/j.molcel.2014.03.034

SUMMARY

CCR4-NOT is a major effector complex in miRNA-mediated gene silencing. It is recruited to miRNA targets through interactions with tryptophan (W)-containing motifs in TNRC6/GW182 proteins and is required for both translational repression and degradation of miRNA targets. Here, we elucidate the structural basis for the repressive activity of CCR4-NOT and its interaction with TNRC6/GW182s. We show that the conserved CNOT9 subunit attaches to a domain of unknown function (DUF3819) in the CNOT1 scaffold. The resulting complex provides binding sites for TNRC6/GW182, and its crystal structure reveals tandem W-binding pockets located in CNOT9. We further show that the CNOT1 MIF4G domain interacts with the C-terminal RecA domain of DDX6, a translational repressor and decapping activator. The crystal structure of this complex demonstrates striking similarity to the eIF4G-eIF4A complex. Together, our data provide the missing physical links in a molecular pathway that connects miRNA target recognition with translational repression, deadenylation, and decapping.

INTRODUCTION

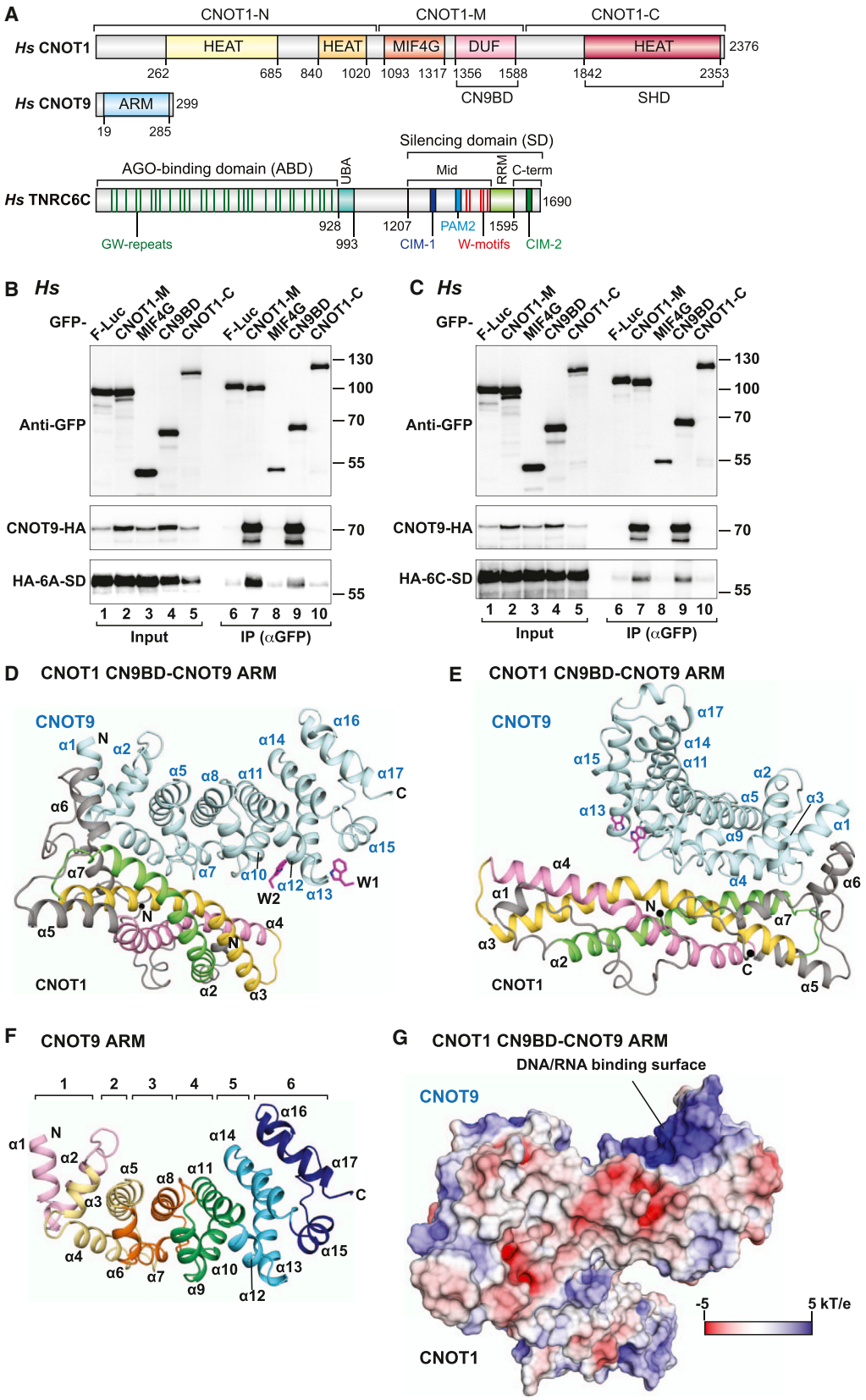
miRNAs are endogenous noncoding RNAs that associate with Argonaute proteins (AGOs) into miRNA-induced silencing complexes (miRISCs) and posttranscriptionally silence the expression of mRNAs containing complementary sequences (Ameres and Zamore, 2013). In animals, most mRNA targets are only partially complementary to the miRNA. As a result, even catalytically active AGOs cannot cleave the mRNA target and recruit additional proteins to mediate silencing (Fabian and Sonenberg, 2012; Braun et al., 2013).

GW182 proteins are the best-characterized AGO-binding partners required for miRNA-mediated silencing in animal cells (Fabian and Sonenberg, 2012; Braun et al., 2013). There are three GW182 paralogs (termed TNRC6A, TNRC6B, and TNRC6C) in vertebrates and only one family member in *Drosophila melanogaster* (*Dm* GW182; Braun et al., 2013). These proteins bind directly to AGOs via multiple GW (glycine and tryptophan) repeats present in the N-terminal AGO-binding domain and then recruit effector complexes, such as the CCR4-NOT and PAN2-PAN3 deadenylase complexes (Braun et al., 2013).

Similar to the interaction with AGOs, GW182 proteins interact with the deadenylase complexes through W-containing motifs, which in this case are located in their C-terminal silencing domains (SDs) (Figure 1A; Braun et al., 2011; Chekulaeva et al., 2011; Christie et al., 2013; Fabian et al., 2011; Huntzinger et al., 2013). The crystal structures of human AGO2 and *Dm* PAN3 homodimers have revealed the molecular basis for their interaction with the W-containing motifs in GW182 proteins (Schirle and MacRae, 2012; Christie et al., 2013). However, it is not known how the W motifs interact with the CCR4-NOT complex.

The CCR4-NOT complex consists of several independent modules that dock with the NOT1 subunit, which serves as a scaffold for complex assembly (Wahle and Winkler, 2013). NOT1 consists of an N-terminal (NOT1-N), middle (NOT1-M), and C-terminal (NOT1-C) region (Figures 1A and S1A available online). The NOT1-M region comprises two domains: a MIF4G domain that is structurally related to the middle domain of eIF4G and a domain of unknown function (DUF3819). The MIF4G domain interacts with the catalytic module of the CCR4-NOT complex, comprising the two deadenylases CAF1 (or its paralog POP2) and CCR4a (or its paralog CCR4b) (Wahle and Winkler, 2013). The DUF3819 domain interacts with the highly conserved Armadillo (ARM) repeat domain of the CNOT9 subunit (Figure 1A; also known as CAF40, RQCD1, or RCD1) (Bawankar et al., 2013). Importantly, the CNOT1-M region has been shown to mediate the interaction of CCR4-NOT with the silencing domain of TNRC6C (6C-SD) (Huntzinger et al., 2013). Furthermore, the RNA helicase DDX6 (also known as Dhh1, RCK/p54, or Me31B) also interacts with CNOT1 (Coller et al., 2001; Hata et al., 1998; Maillet and Collart, 2002). DDX6 functions as a translational repressor and decapping activator (Presnyak and Coller, 2013) and has previously been implicated in the miRNA pathway (Chu and Rana, 2006; Eulalio et al., 2007; Nishihara et al., 2013).

The CCR4-NOT complex not only mediates mRNA deadenylation, but it is also required both for the translational



(legend on next page)

repression and degradation of miRNA targets (Braun et al., 2011; Chekulaeva et al., 2011; Fabian et al., 2011; Huntzinger et al., 2013; Zekri et al., 2013). miRNA target degradation is catalyzed by the enzymes of the 5′-to-3′ mRNA decay pathway (Huntzinger and Izaurralde, 2011). In this pathway, mRNAs are first deadenylated by the CCR4-NOT complex, decapped by the decapping enzyme DCP2 and additional coactivators, and finally degraded from the 5′ end by the exonuclease XRN1. However, it has remained unclear whether and how decay factors, including the decapping enzyme DCP2 and the 5′ exonuclease XRN1, would be recruited directly by the CCR4-NOT complex. Similarly, the mechanism of translational repression by CCR4-NOT has remained elusive, although it has recently been proposed that the DEAD-box protein eIF4A2 may play a role (Meijer et al., 2013). In particular, it was suggested that the CNOT1 MIF4G domain could recruit eIF4A2 via an interaction similar to the one observed in the eIF4G-eIF4A complex (Meijer et al., 2013). eIF4A2 could then prevent ribosome scanning by clamping onto the 5′ UTR of the mRNA (Meijer et al., 2013).

In this study, we investigated the molecular basis for the interactions mediated by the human CNOT1-M region and its role in silencing. We show that the CNOT1 DUF3819 domain forms a binary complex with the ARM repeat domain of the CNOT9 subunit and was hence termed the CN9BD (*CAF40/CNOT9 b*inding *d*omain). We pinpoint the CN9BD-CNOT9 complex as a primary binding site for the TNRC6 silencing domains (TNRC6-SDs) in the CNOT1-M region. The crystal structure of the CN9BD-CNOT9 complex shows that the CN9BD adopts a defined fold that is recognized and stabilized by CNOT9 binding. This structure also reveals tandem W-binding pockets in CNOT9, suggesting a mechanism for the recruitment of the CCR4-NOT complex by W-rich GW182/TNRC6 proteins. Furthermore, we identified the CNOT1 MIF4G domain as a specific binding partner for DDX6 rather than eIF4A2. The crystal structure of the MIF4G-DDX6 complex reveals an eIF4G-eIF4A-like interaction, which is compatible with the simultaneous binding of CAF1 to the MIF4G domain and the recruitment of decapping factors (EDC3, LSm14A, or Pat) by DDX6. Together, our data now allow us to trace the direct molecular interactions that ultimately lead from miRNA target recognition to the translational repression, deadenylation, and decapping of the mRNA target.

RESULTS

Dual Functions of the CNOT1 DUF3819 Domain in TNRC6 and CNOT9 Binding

To pinpoint the TNRC6-SD binding site within the CNOT1-M region (Figures 1A and S1A), we performed coimmunoprecipitation assays in human embryonic kidney 293T (HEK293T) cells. We found that TNRC6A-SD and TNRC6C-SD interacted primarily with the DUF3819 domain rather than with the MIF4G domain (Figures 1B and 1C, lane 9), overlapping with the sequences required to bind the ARM repeat domain of CAF40/CNOT9 in human cells (Figures 1B and 1C, lane 9, and S1B) and in *Dm* Schneider (S2) cells (Figure S1C; Bawankar et al., 2013). The DUF3819 domain was therefore termed the CAF40/CNOT9 binding domain or CN9BD.

To elucidate the molecular basis for the observed interactions, we expressed suitable protein constructs in *Escherichia coli* (Table S1) and used specific tags for pull-down assays (Figure S1B) and protein purification. The ARM repeat domain of CNOT9 (residues 19–285) formed homodimers in solution, as observed previously (Garces et al., 2007) and as confirmed by size-exclusion chromatography and multiangle static laser light scattering (MALLS; Figure S1D). Protein constructs corresponding to the CN9BD precipitated upon removal of the solubility tag, indicating instability or improper folding in the absence of CNOT9. However, several CN9BD constructs (residues 1,356–1,581, 1,356–1,607, and 1,356–1,628) could be coexpressed and copurified with the ARM repeat domain of CNOT9 (Figure S1B). Under these conditions, size-exclusion chromatography and MALLS indicate the formation of a binary CN9BD-CNOT9 complex that appears to be mutually exclusive with CNOT9 homodimerization (Figures S1D and S1E).

We obtained crystals of the CN9BD-CNOT9 complex that diffracted X-rays to a 1.65 Å resolution (Table 1). Crystals were exclusively obtained using the CN9BD 1,356–1,607 construct, although a shorter construct (CN9BD 1,356–1,581) was also soluble and formed a binary complex with CNOT9 (Figure S1B). In a bid to identify putative TNRC6 binding sites, we also crystallized the complex in the presence of L-tryptophan (W). These crystals diffracted X-rays to a 2.05 Å resolution (Table 1).

Figure 1. Structure of the CNOT1 CN9BD Bound to the CNOT9 ARM Repeat Domain

(A) Diagram of CNOT1 with N-terminal, middle, and C-terminal regions (CNOT1-N, CNOT1-M, and CNOT1-C, respectively). CNOT1-N consists of two HEAT-like repeat domains. CNOT1-M contains an MIF4G domain and the CN9BD (previously DUF3819). CNOT1-C contains the NOT1 superfamily homology domain (SHD). CNOT9 contains an armadillo (ARM) repeat domain. TNRC6C contains and N-terminal AGO-binding domain (ABD), a ubiquitin-associated-like domain (UBA), and a C-terminal silencing domain (SD). The SD comprises a Mid region, an RNA recognition motif (RRM), and a C-terminal (C-term) region. The positions of the CCR4-NOT interacting motifs 1 and 2 (CIM-1 and CIM-2) and the PAM2 motif (PABP-interacting motif 2) are indicated. Vertical green and red lines indicate the positions of W-containing motifs binding to AGOs and deadenylases, respectively. Amino acid positions at domain boundaries are indicated below the protein outlines. See also Figure S1A.

(B and C) Western blot showing the interaction between GFP-CNOT1 fragments and HA-MBP-tagged CNOT9 and HA-tagged TNRC6A-SD (6A-SD) or TNRC6C-SD (6C-SD) in HEK293T cells. GFP-F-Luc served as a negative control. See also Figures S1B–S1H.

(D and E) Overall structure of the CN9BD-CNOT9 binary complex in two orientations. CNOT9 is shown in cyan; the three long α helices in the CNOT1 CN9BD are shown in pink, yellow, and green. Bound W residues are shown as magenta sticks. Secondary structure elements are labeled in blue for CNOT9 and in black for CNOT1.

(F) Cartoon representation of the ARM repeat domain of CNOT9. The ARM repeats are colored in a gradient from pink to dark blue from the N to C terminus.

(G) Surface representation of the CN9BD-CNOT9 binary complex in the same orientation as that in (D) and colored according to surface potential contoured from −5 kT/e (red) to +5 kT/e (blue). The position of the DNA/RNA binding surface is indicated. See also Figures S1F and S1G.

Table 1. Data Collection and Refinement Statistics

Data Set	CNOT1-CNOT9 Complex	CNOT1-CNOT9-W Complex	CNOT1-DDX6 Complex
Space group	C2	C2	$P2_1\,2_1\,2_1$
Unit Cell			
Dimensions (a / b / c) (Å)	154.8 / 67.2 / 72.3	154.1 / 66.8 / 72.0	43.9 / 90.8 / 95.8
Angles (α / β / γ) (°)	90 / 99.6 / 90	90 / 100.3 / 90	90 / 90 / 90
Data Collection			
Wavelength (Å)	1.000	1.000	1.000
Resolution range (Å)	76.3–1.65 (1.69–1.65)	61.1–2.05 (2.10–2.05)	65.9–1.75 (1.80–1.75)
R_{sym} (%)	2.9 (63.6)	6.0 (78.6)	5.3 (78.9)
Completeness (%)	99.5 (99.0)	99.5 (99.0)	99.2 (99.5)
Mean $I/\sigma(I)$	19.8 (1.8)	14.9 (1.8)	17.1 (2.3)
No. of unique reflections	87,658 (6,377)	45,149 (3,317)	39,120 (2,837)
Multiplicity	3.4 (3.4)	6.8 (6.4)	5.0 (5.0)
Refinement			
Data range (Å)	76.3–1.65	61.1–2.05	47.9–1.75
R_{work} (%)	16.0	17.6	16.2
R_{free} (%)	17.9	21.5	20.3
No. of Atoms per Asymmetric Unit			
All atoms (no H)	4,590	4,178	3,386
Nonsolvent	4,116	4,030	3,193
Water	474	148	193
Average B Factor (Å^2)			
All atoms	38.5	64.3	31.3
Nonsolvent	37.9	64.5	30.7
Water	43.6	58.7	41.3
Ramachandran Plot			
Favored regions (%)	99.0	98.8	99
Disallowed regions (%)	0	0	0
Root-Mean-Square Deviation from Ideal Geometry			
Bond lengths (Å)	0.011	0.006	0.011
Bond angles (°)	1.27	0.88	1.31

Values in parentheses are for highest-resolution shell.

Crystal Structures of the CN9BD-CNOT9 Complex

The structures were solved by molecular replacement using the CNOT9 structure (Protein Data Bank ID code [PDB] 2FV2; Garces et al., 2007) as a search model and refined to final R_{free} values of 17.9% in the absence and 21.5% in the presence of L-tryptophan (Table 1 and Figures 1D and 1E). Overall, the two structures are highly similar and superpose with a root-mean-square deviation (rmsd) of 0.16 Å over 431 Cα atoms. The CN9BD-CNOT9 complex adopts a V-shaped architecture resulting from the rod-shaped CN9BD bound to the N-terminal, convex surface of the crescent-shaped CNOT9 (Figure 1E).

CNOT9 consists of 17 α helices arranged into 6 imperfect ARM repeats (Figures 1F and S2). The structure of CNOT9 in the complex is virtually identical to the structure of the isolated CNOT9 ARM repeat domain (rmsd of 0.51 Å over 246 equivalent Cα positions; Figure S1F; Garces et al., 2007), indicating that CNOT9 does not undergo major structural rearrangements upon binding to CNOT1. The CN9BD-CNOT9 interface partially overlaps with the CNOT9 homodimerization interface, as determined from the structure of the isolated CNOT9 (Figure 1D versus S1G and Figures S3A–S3D; Garces et al., 2007). This observation explains why CNOT1 binding is mutually exclusive with CNOT9 homodimerization. In both of these alternative complexes, a positively charged cleft located on the opposite, concave face of the CNOT9 crescent remains exposed and fully accessible to interact with potential RNA or DNA ligands (Figure 1G; Garces et al., 2007).

Intriguingly, the C-terminal tail of the CN9BD (residues 1,589–1,607) is largely disordered but required for crystallization. It makes a crystal contact with a neighboring CNOT9 molecule, inserting W1603 into a specific W-binding pocket (pocket 1). Consequently, W1603 is clearly defined in both structures (Figures 1D and 1E; W1); at 1.65 Å resolution, the backbone of the flanking residues 1,601–1,604 is also visible (see below). A second W-binding pocket (pocket 2) on CNOT9 becomes apparent and occupied in the presence of additional L-tryptophan in the crystallization solution (Figures 1D and 1E; W2).

CN9BD Folds into a Defined Domain Arranged as a Three-Helix Bundle

In contrast to all CNOT1 subdomains structurally determined so far, the CN9BD is not organized into HEAT-like repeats. Instead, it is composed of seven α helices, with helices α2–α4 arranged as an antiparallel, rod-shaped bundle of three kinked α helices (Figures 1D and 1E). These helices comprise 42–44 residues and are kinked once (α3) or twice (α2 and α4). The side of the bundle that binds CNOT9 is flanked by three additional short helices (helices α5–α7; Figures 1E and S4). Two of these helices (α6 and α7), together with the connecting loop L6, are part of the interaction interface and are likely adaptable. In the crystal, their position might also be influenced by packing interactions because loop L6 and helix α6 contact the equivalent structural elements of a neighboring symmetry mate (Figure S4B). However, a CNOT1 fragment lacking helix α7 (1,356–1,561) did not interact with CNOT9 in solution (Figure S1B), indicating that these helices and the spatial arrangement observed in the crystal are indeed relevant for complex formation.

One remarkable structural feature of the CN9BD is that both the N- and C-terminal peptides fold back on the three-helix bundle, forming multiple hydrophobic interactions (Figures 2A, 2B, and S4C). As a result, the N and C termini of the domain enter close proximity and form a main-chain contact between Y1357 and F1582 (Figures 2A and 2B). Furthermore, the C-terminal peptide (around P1580-G1581) becomes locked into its conformation by interactions with CNOT9 residues V71 and P75 (Figure 2C), consistent with the observation that a deletion of the C-terminal peptide destabilizes the interaction between the *Dm* proteins (Figure S1C). Overall, the CN9BD adopts a rather unique protein fold that has no close structural homologs in the PDB.

CNOT9 Recognizes and Stabilizes the Fold of the CN9BD

The interface between CNOT1 and CNOT9 is larger than the interface for CNOT9 homodimerization (3,170 Å^2 versus 2,005 Å^2; Figures S3B–S3D) and thus likely favored in vivo. It is predominantly hydrophobic and centered on the first three N-terminal ARM repeats (ARM1–ARM3) of CNOT9, which engage a surface composed of helices α2, α3, α6, α7, loop L6, and the very C-terminal end (C-term) of the CNOT1 CN9BD (Figures 1D, 1E, 2A, 2C, and 2D).

ARM1 includes an important serine (S30) and fixes loop L6 and helix α7, whereas ARM3, which includes an important phenylalanine (F118), contacts helices α2 and α3 (Figure S2). ARM2 makes numerous contacts, including each of the CNOT1 interface elements. In particular, the central F60 in the elbow loop L3, between ARM2 helices α3 and α4, and the following helix α4 insert into a hydrophobic groove of the CN9BD that is flanked by CNOT1 helices α2, α3, and α7 (Figures 2A–2D, S2, and S4). The respective interface residues include F60 (L3), G61 (L3), A64 (α4), L67 (α4), V71 (α4), Y74 (L4), P75 (L4), and P79 (L4) on the side of CNOT9. On the side of CNOT1, there are T1419 (α2), I1423 (α2), K1426 (α2), D1427 (α2), F1428 (α2), M1444 (α3), L1448 (α3), G1451 (α3), I1455 (α3), M1452 (α3), R1458 (α3), L1559 (L6), V1564 (L6), Q1568 (L6), L1569 (L6), V1571 (α7), Y1572 (α7), F1575 (α7), P1580 (C-term), and G1581 (C-term). These complex hydrophobic contacts allow CNOT9 to recognize and stabilize the CN9BD as a folded domain.

Additional polar flanking interactions provide further specificity and, in particular, allow helix α6 of the CN9BD to adapt to the surface of CNOT9. Here, CNOT9 H58 plays an important role. It is read out by water-mediated hydrogen bonds from Q1549 (α6) and from the invariant lysine K1426 (α2) of the CN9BD and additionally stacks on Y1548 (α6; Figure 2D).

To validate the interface, we initially generated single point mutations, but these only reduced the interaction (e.g., F60A) or had no effect (A64Y; Figure S4D). To disrupt the interface, it was necessary to introduce four mutations in CNOT9 (Mut1; H58A, F60A, A64Y, V71Y) or in the CNOT1 CN9BD (4×M; K1426S, G1451Y, R1458A, Q1549A; Figures S4D and S4E). Importantly, the CNOT9 Mut1 retained the ability to homodimerize (Figure S1H), indicating that the mutations do not disrupt protein folding.

For in vivo coimmunoprecipitation assays in HEK293T cells, it was also necessary to use the quadruple CNOT9 mutation (Mut1) to disrupt the interaction with full-length CNOT1 or with the CNOT1-M region (Figures 2E and S4F). In CNOT1, it was even necessary to generate a quintuple mutation (5×M; I1423D plus 4×M) because the quadruple mutation still showed residual binding (Figures 2F and S4G). These results are consistent with a rather extensive and high-affinity interaction.

Finally, to demonstrate the conservation of the interface, we repeated the coimmunoprecipitation experiments in *Dm* S2 cells. Notably, double mutations in CNOT9 and single G1562E or K1537S mutations in CNOT1 (corresponding to human G1451 and K1426) were sufficient to disrupt the interaction in this case (Figures 2G, 2H, and S4H). These results suggest that the interface is indeed conserved, but that affinities may vary among orthologs.

CNOT9 Harbors Tandem Tryptophan-Binding Pockets

Perhaps one of the most striking features of the CNOT1-CNOT9 structure is the presence of tryptophan residues bound to tandem hydrophobic pockets in CNOT9. These pockets are located on ARM5, on either side of the kink between helices α12 and α13, on the convex surface of the crescent (Figures 3 and S2).

Pocket 1 is flanked by helix α15 of ARM6 and lined by residues C200, Q201, T202, Y203, F206, R244, A245, and A248 (Figures 3A–3C). It accommodates W1603 (W1) from a symmetry-related CNOT1 molecule in both crystal structures. W1603 stacks between Y203 and R244 and is specifically recognized on its N7 nitrogen by the carbonyl group of C200 (Figures 3A–3C). Pocket 2 is flanked by helix α10 of ARM4 and lined by residues I164, P165, L168, Y198, I199, R205, H208, and V209. It accommodates a free L-tryptophan (W2) that is absent in the high-resolution structure obtained in the absence of L-tryptophan in the crystallization condition. W2 stacks between P165 and R205 and is also recognized specifically on its N7 nitrogen by a hydrogen bond to H208 (Figures 3B, 3D, and S2).

Similar to the situation in the AGO2 structure, the distance between the two bound tryptophan residues is 20–25 Å. This is a typical intervening distance between tryptophan residues in TNRC6 proteins and can be spanned by approximately 8–10 residues in an extended conformation (Schirle and MacRae, 2012).

The CNOT9 W-Binding Pockets Are Binding Sites for GW182/TNRC6 Proteins

To test whether the W-binding pockets in CNOT9 represent bona fide binding sites for TNRC6 proteins, we substituted residues lining the pockets and performed pull-down assays in vitro. In pocket 1, we substituted A248 with phenylalanine to fill the cavity with a bulky side chain. We also substituted Y203 and R244 with alanine to prevent stacking interactions. In pocket 2, we substituted P165 with glycine and R205 with alanine to prevent stacking interactions and H208 with alanine to prevent specific recognition of the W ligand via hydrogen bonding (Table S1).

We observed that the TNRC6A-SD (6A-SD), which was expressed in *E. coli* with a maltose-binding protein (MBP) tag, pulled down CNOT9 in the absence of CNOT1, indicating that the interaction is direct (Figures 4A, lane 7, and 4B). Furthermore, efficiently disrupting this interaction required mutating both W-binding pockets simultaneously, demonstrating that they are indeed responsible for binding TNRC6 SDs (Figure 4A, lanes 10–12 as compared to lanes 8 and 9). Similar results were obtained for the TNRC6C-SD (Figure S5A). Importantly, the mutations in the W-binding pockets did not affect CNOT9 homodimerization (as shown by MALLS; Figures S5B–S5G) or binding to the CNOT1 CN9BD (Figure S5H), indicating that the mutations do not disrupt the CNOT9 fold.

Finally, a complex containing the CN9BD bound to the CNOT9 P1+P2a mutant did not interact with the 6A-SD (Figure S5I), indicating that the W-binding pockets in CNOT9 represent the primary binding sites for TNRC6 proteins in the CN9BD-CNOT9 complex and that the CN9BD is not sufficient to recruit TNRC6 proteins.

Next, we investigated the contribution of the CNOT9 W-binding pockets to TNRC6 binding in a cellular context. To this end, the mutations described above were introduced into full-length

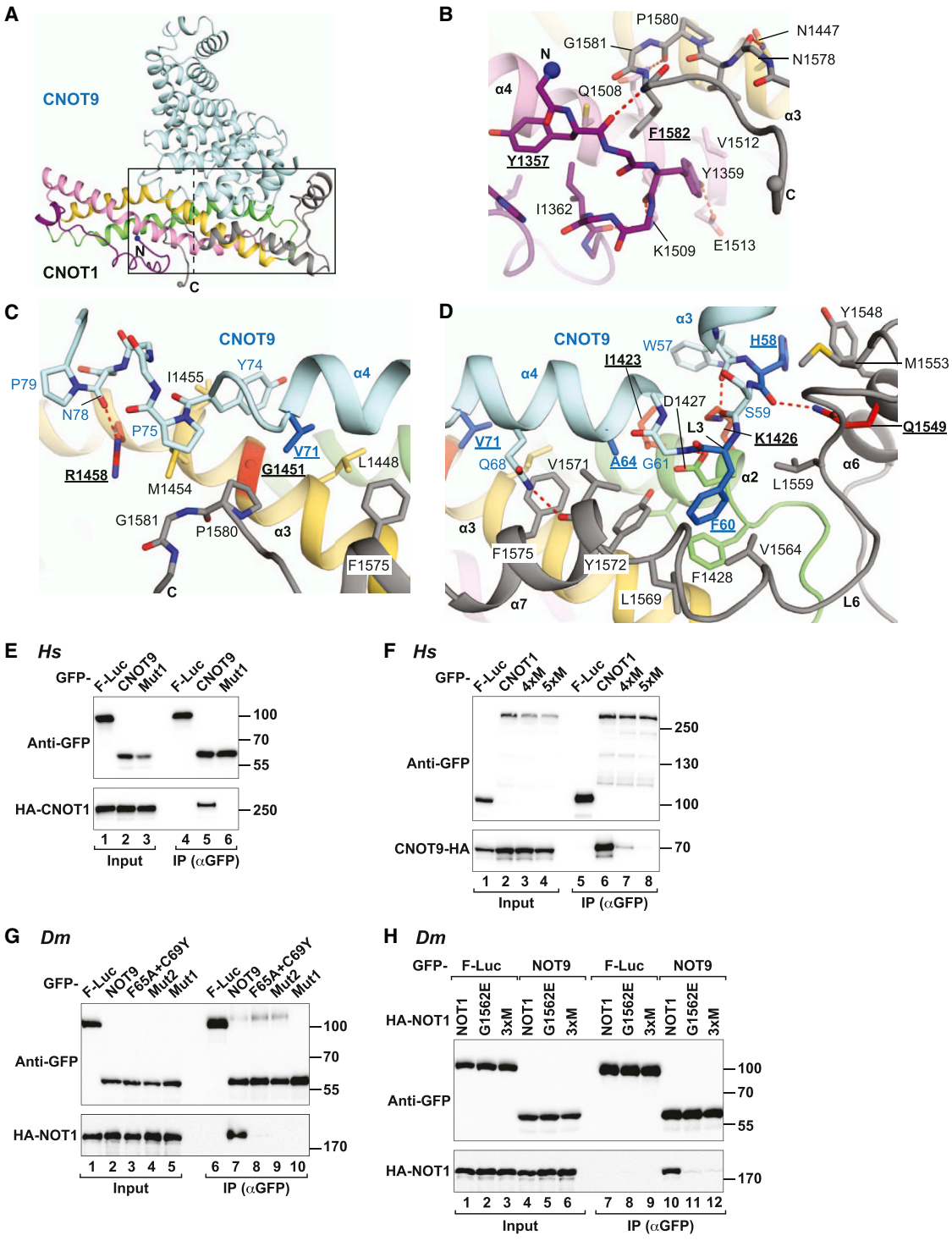


Figure 2. The CN9BD-CNOT9 Binding Interface

(A–D) Overview and close-up views of the interface between the CNOT1 CN9BD and the CNOT9 ARM domain, with selected interface residues shown as sticks. Residues mutated in this study are underlined and shown as red sticks (CNOT1) or dark blue sticks (CNOT9). The small and large rectangles in (A) indicate the views shown in (C) and (D), respectively. Residues and secondary structure elements are labeled in black for CNOT1 and in blue for CNOT9. Hydrogen bonds are shown as red dashed lines. See also Figures S2–S4.

(E) Interaction of GFP-CNOT9 (full-length wild-type or quadruple mutant, Mut1) with HA-CNOT1 (full length) in human cells. See also Figure S4.

(legend continued on next page)

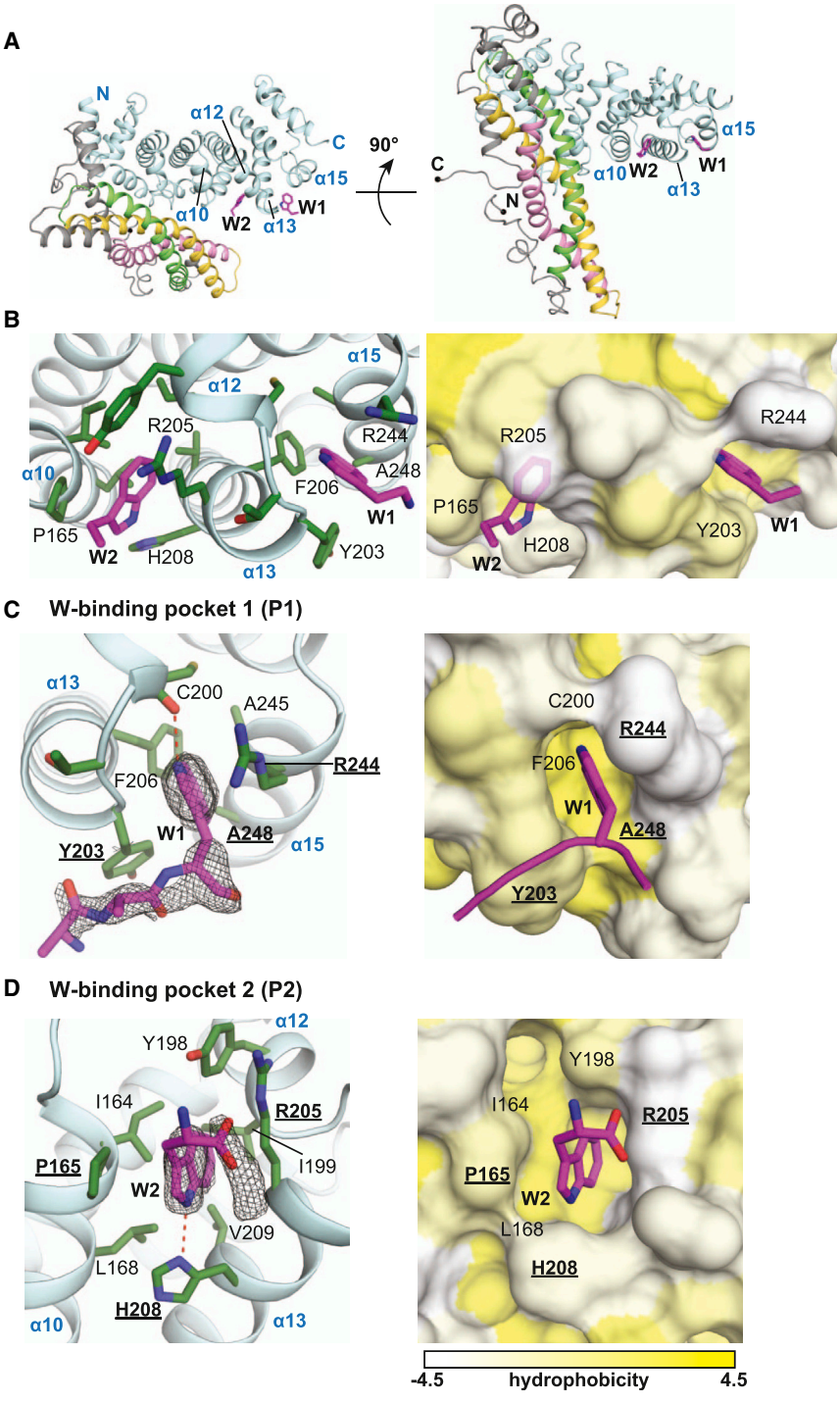


Figure 3. CNOT9 Harbors Tandem W-Binding Pockets

(A) Overview of the CN9BD-CNOT9 complex is shown for orientation. The W residues (W1 and W2 corresponding to CNOT1 W1603 and free W, respectively) bound to CNOT9 are shown as sticks. (B) Close-up views of the W-binding pockets. The right panel shows a surface representation of the view in the left panel, with the CNOT9 surface colored white to yellow with increasing hydrophobicity (scores according to Kawashima et al., 2008). (C and D) Close-up views of W-binding pocket 1 (C) (high-resolution structure) and W-binding pocket 2 (D) (low-resolution structure). The electron difference densities (Fo-Fc, contoured at 2.5 σ) for the W-containing peptide and the free W residue are shown as a gray mesh, and the corresponding structural models are displayed as magenta sticks. Residues mutated in this study are underlined. Hydrogen bonds are shown as red dashed lines. See also Figures S2 and S4 and Table S1.

CNOT9 and tested for interaction with TNRC6s. The CN9BD-CNOT9 complex interacts with the 6A-SD and 6C-SD (Figures 4C and S5J, lane 8). These interactions were abolished by mutations in the CNOT9 W-binding pockets (Figures 4C and S5J, lanes 9–11). As expected, the pocket mutations did not interfere with CNOT1 binding (Figures 4C, 4D, and S5J). Surprisingly, CNOT9 Mut1, which cannot interact with CNOT1, also failed to interact with the TNRC6 proteins (Figures 4C and S5J, lane 12), although this mutant interacts with the 6A-SD in vitro (Figure 4B, lane 12). These observations suggest that, in vivo, CNOT9 only binds TNRC6s when it is assembled into the CCR4-NOT complex.

Importantly, CNOT1 (or the CCR4-NOT complex) provides binding sites for TNRC6s independently of CNOT9 because full-length CNOT1 mutants (4×M and 5×M) that are defective in binding CNOT9 (Figure 2F) still interact with 6C-SD and 6A-SD (Figures 4E and S5K).

To investigate whether the role of the W-pockets in NOT9 is conserved, we performed similar coimmunoprecipitation assays with the *Dm* NOT9 and GW182 proteins in *Dm* S2 cells. In agreement with the results in human cells, we made the following observations. First, the NOT9 W-binding pockets are the major GW182-binding sites in the CN9BD-NOT9 complex because mutations in these pockets abrogate the interaction with GW182 without affecting complex formation (Figure 4F, lanes 8–11). Second, NOT9 interacts with GW182

(F) Interaction of GFP-CNOT1 (full-length wild-type or mutants) with CNOT9-HA-MBP in HEK293T cells.
(G) Interaction of GFP-tagged *Dm* NOT9 (wild-type or mutants) with HA-NOT1 in *Dm* S2 cells.
(H) Interaction of GFP-tagged *Dm* NOT9 with HA-NOT1 (wild-type or mutants) in *Dm* S2 cells. In all panels, cell lysates were treated with RNase A prior to immunoprecipitation. See also Figure S4 and Table S1.

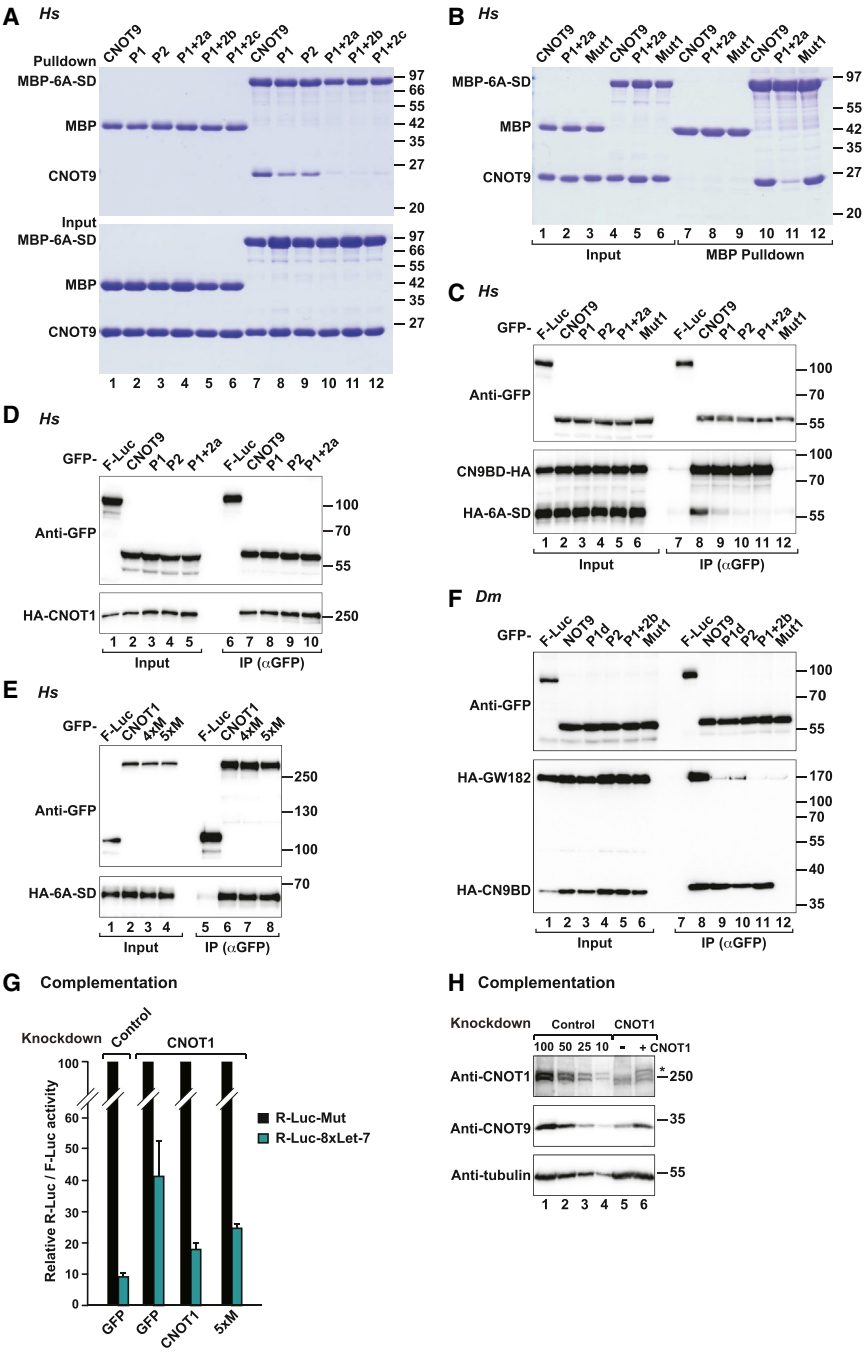


Figure 4. The CNOT9 W-Binding Pockets Mediate Binding to the TNRC6s

(A) MBP pull-downs using recombinant MBP-tagged 6A-SD and His-tagged CNOT9 ARM domain (wild-type or the indicated mutants). MBP served as a negative control. See also Figure S5.

(B) MBP pull-downs using recombinant MBP-tagged 6A-SD and His-tagged CNOT9 ARM domain (wild-type or the indicated mutants).

(C) Interaction between GFP-CNOT9 (wild-type or the indicated mutants) and HA-6A-SD in the presence of HA-MBP-tagged CN9BD in HEK293T cells.

(D) Interaction between GFP-CNOT9 (wild-type or the indicated mutants) and full-length HA-CNOT1 in HEK293T cells.

(E) Interaction between GFP-CNOT1 (wild-type or the 4×M and 5×M mutants that do not bind CNOT9) and HA-6A-SD in HEK293T cells.

(F) Interaction between GFP-CNOT9 (wild-type or the indicated mutants) and HA-GW182 in the presence of HA-tagged CN9BD in *Dm* S2 cells. See also Figure S5 and Table S1.

(G) HeLa cells (transfected with a control shRNA) or cells depleted of CNOT1 were transfected with a mixture of three plasmids: the psiCHECK-8×Let-7 or the corresponding reporter carrying mutations in Let-7-binding sites (R-Luc-Mut), a plasmid expressing F-Luc as a transfection control, and a plasmid expressing shRNA-resistant versions of GFP-CNOT1 (wild-type or 5×M mutant) or GFP. For each condition, *Renilla* luciferase activity was measured, normalized to that of the F-Luc transfection control, and set at 100% in cells expressing R-Luc-Mut (black bars). Mean values ± SD from five independent experiments are shown.

(H) Western blots showing the efficiency of the CNOT1 knockdown and the expression levels of endogenous CNOT9. Dilutions of control cell lysates were loaded in lanes 1–4 to estimate the efficacy of the depletion. α-tubulin served as a loading control. The asterisk indicates the position of the GFP-CNOT1 used in the complementation assay. See also Figures S5N and S5O.

only when bound to NOT1 in cell lysates (Figure 4F, lane 12). Third, an extended M region of NOT1 is sufficient for binding to GW182 proteins, independent of CNOT9 (Figures S5L and S5M, lane 6 versus 5).

Collectively, our results indicate that CNOT1 (or the CCR4-NOT1 complex) provides multiple binding sites for GW182/TNRC6 proteins, with two sites located in the CN9BD-CNOT9 module. The presence of multiple binding sites in CNOT1 is consistent with the observation that the GW182/TNRC6 proteins contact CNOT1 through multiple W-containing motifs that contribute additively to the interaction (Braun et al., 2011; Chekulaeva et al., 2011; Fabian et al., 2011; Huntzinger et al., 2013).

In agreement with the redundancy of binding sites, we observed that the CN9BD-CNOT9 interaction contributes to, but is not essential for, silencing. Indeed, the CNOT1 5×M mutant that does not interact with CNOT9 was still able to rescue silencing of a Let-7 reporter (psiCHECK-8×Let-7; Iwasaki et al., 2009) in cells depleted of endogenous CNOT1, although not as efficiently as wild-type CNOT1 (Figure 4G). Western blot analysis indicated that the levels of CNOT1 in the depleted cells were reduced to 10% of the control levels (Figure 4H). Interestingly, CNOT9 levels were also strongly reduced, suggesting that CNOT9 is destabilized in the absence of CNOT1 (Figure 4H, lane 5). In contrast, the levels of endogenous AGO2 and TNRC6A

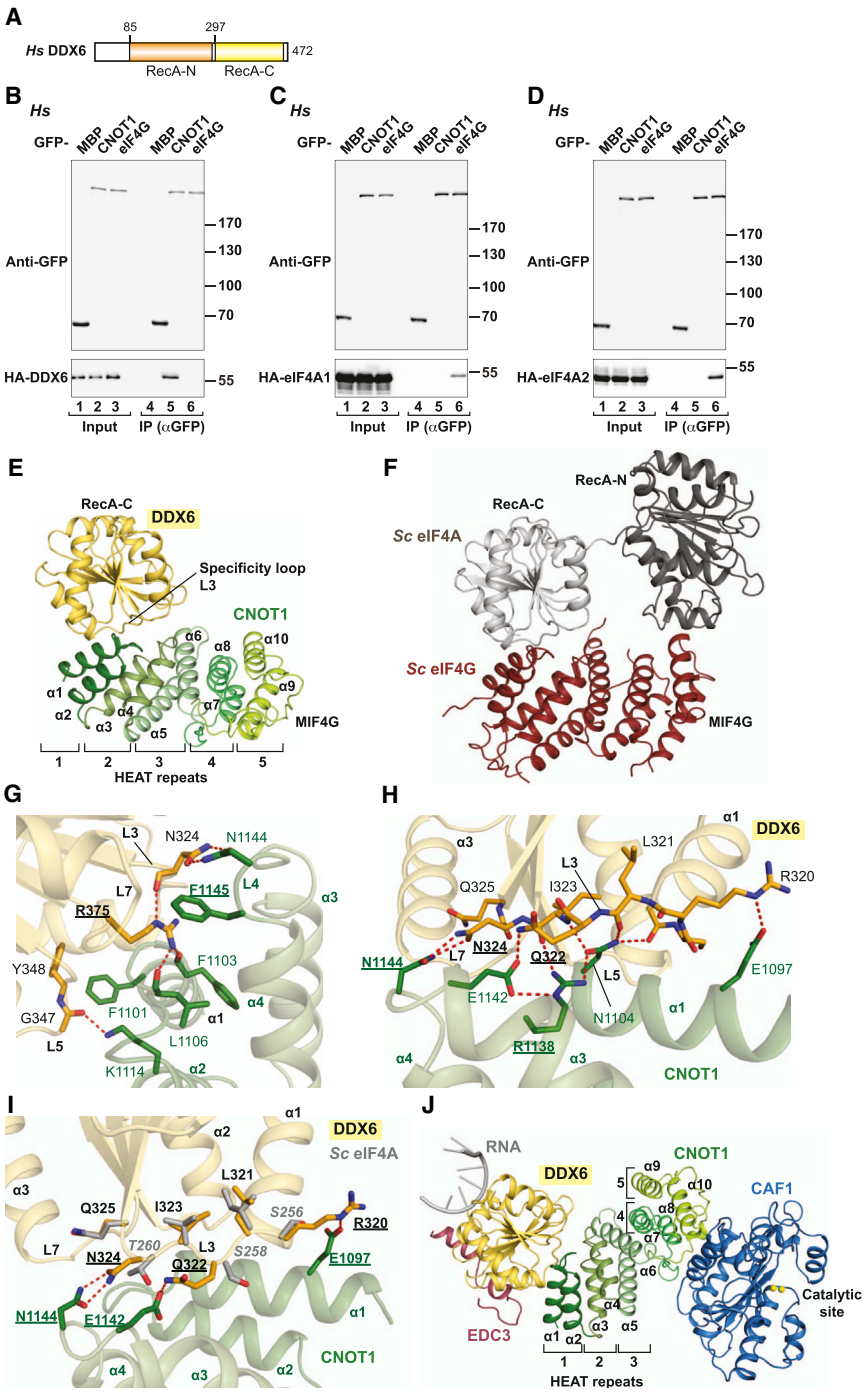


Figure 5. Structure of the CNOT1 MIF4G Domain Bound to the DDX6 RecA-C Domain

(A) DDX6 consists of two RecA-like domains, termed RecA-N and RecA-C, connected by a flexible linker.

(B–D) Interaction of GFP-CNOT1 or GFP-eIF4G with HA-tagged DDX6 (B), eIF4A1 (C), or eIF4A2 (D) in HEK293T cells. See also Figure S6.

(E and F) Overall structure of the CNOT1 MIF4G-DDX6 RecA-C complex (this study) (E) and the *Saccharomyces cerevisiae* eIF4G MIF4G-eIF4A complex (2VSO; Schütz et al., 2008) (F). Selected secondary structure elements are indicated.

(G and H) Close-up views of the interface between the CNOT1 MIF4G domain and the DDX6 RecA-C domain showing the DDX6 arginine anchor residue R375 (G) and loop L3 (H). Selected interface residues are shown as sticks and colored green (CNOT1) or orange (DDX6). Residues and secondary structural elements are labeled in green for CNOT1 and in black for DDX6. Residues mutated in this study are underlined.

(I) Superposition of *Sc* eIF4A loop L3 (residues 255–262) onto *Hs* DDX6 L3 (329–336). Selected interface residues in loop L3 of the DDX6 and eIF4A RecA-C domains are shown as sticks and colored in orange (DDX6) and gray (eIF4A). The CNOT1 MIF4G residues that specifically form hydrogen bonds with DDX6 residues are shown in green. The residues that contribute to the specificity of the interaction are underlined and shown in bold. eIF4A residues are labeled in italics. Backbone cartoons of *Sc* eIF4G and eIF4A are omitted for clarity.

(J) A structural model built by superposition of DDX6 bound to the EDC3 FDF peptide (PDB 2WAX), DDX6 bound to the CNOT1 MIF4G domain (this study), and the CNOT1 MIF4G bound to CAF1 (PDB 4GMJ). The RNA is modeled based on the structure of Vasa bound to RNA (PDB 2DB3). See also Figure S6J.

were not affected (Figure S5N). The CNOT1 proteins were expressed at comparable levels (Figure S5O). We conclude that the W-binding pockets in CNOT9 assist in the recruitment of the CCR4-NOT complex to miRNA targets.

The MIF4G Domain of CNOT1 Interacts with DDX6

To investigate the hypothesis that the CNOT1 MIF4G domain could directly recruit eIF4A2, we performed coimmunoprecipitation assays using HEK293T cell lysates treated with ribonuclease A (RNase A). We included eIF4A1 as well as DDX6 (Figure 5A) because this helicase has previously been shown to interact with the CCR4-NOT complex in yeast and *Dm* cells (Coller et al., 2001; Hata et al., 1998; Maillet and Collart, 2002; Temme et al., 2010).

We could not observe an interaction between the CNOT1 MIF4G domain and either eIF4A2 or eIF4A1, but we obtained a clear signal with DDX6 (Figures 5B–5D, lane 5). Conversely, eIF4G interacted with both eIF4A1 and eIF4A2 as expected (Yoder-Hill et al., 1993) but exhibited no affinity for DDX6 (Figures 5B–5D, lane 6), indicating that the respective interactions are specific. Further studies indicated that the C-terminal RecA (RecA-C) domain of DDX6 interacted with the CNOT1 MIF4G domain and that this interaction was direct (Figures S6A–S6D).

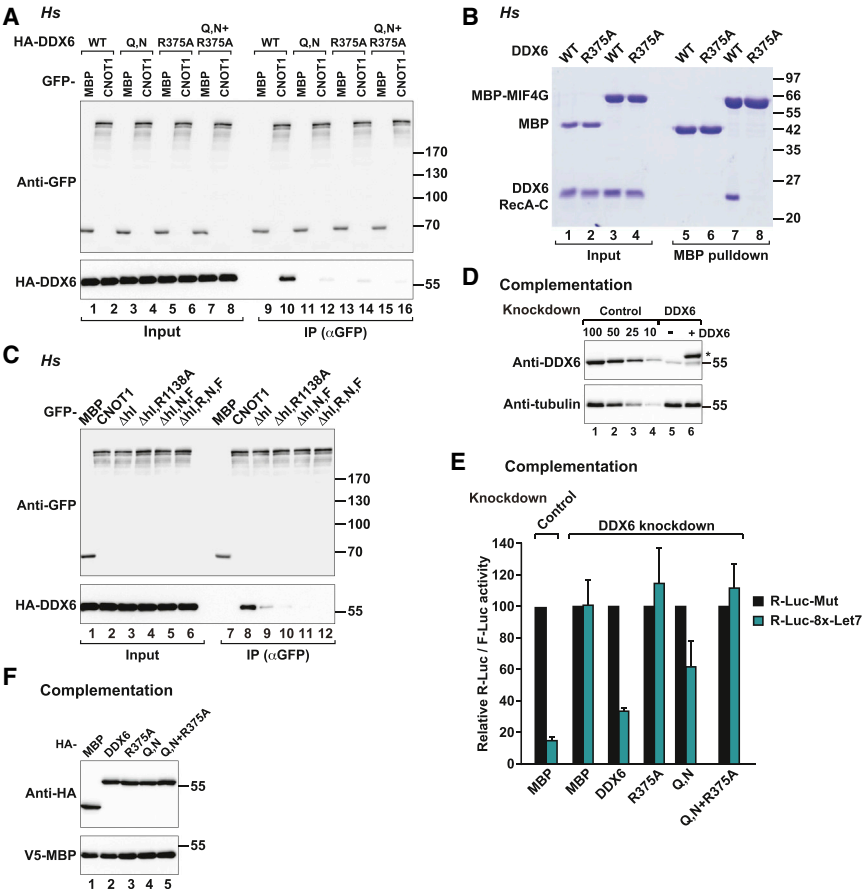


Figure 6. Validation of the CNOT1 MIF4G-DDX6 Binding Interface

(A) Interaction between HA-DDX6 (wild-type or the indicated mutants) and GFP-CNOT1 in HEK293T cells. GFP-MBP served as negative control.

(B) MBP pull-downs using recombinant MBP-tagged CNOT1 MIF4G domain and His-tagged DDX6 RecA-C (wild-type or the R375A mutant). MBP served as a negative control.

(C) Interaction between GFP-CNOT1 (wild-type or the indicated mutants) and HA-DDX6 in HEK293T cells. See also Figure S6 and Table S1.

(D) Western blots showing the efficiency of the DDX6 knockdown. Dilutions of control cell lysates were loaded in lanes 1–4 to estimate the efficacy of the depletion. α-tubulin served as a loading control. The asterisk indicates the position of the HA-DDX6 used in the complementation assay shown in (E).

(E) A complementation assay was performed as described in Figure 4G in cells depleted of DDX6. Mean values ± SD from three independent experiments are shown. See also Figures S5N, S6E, and S6H.

(F) Western blot analysis showing the equivalent expression of the DDX6 proteins used in the complementation assay.

Crystal Structure of CNOT1 MIF4G-DDX6 RecA-C Complex

To understand the molecular basis for the CNOT1-DDX6 interaction, we cocrystallized the CNOT1 MIF4G domain with the DDX6 RecA-C domain and determined the structure of the complex at a 1.75 Å resolution with a final R_{free} of 20.3% (Table 1). The model contains all residues of the CNOT1 MIF4G domain, as well as all residues from the DDX6 RecA-C domain, with the exception of the DDX6 C-terminal residues (454–472). The structures of the two domains from the complex superpose very well with the previously reported structures of the MIF4G domain in isolation (rmsd 0.34 Å; Petit et al., 2012) and the DDX6 RecA-C domain bound to EDC3 (rmsd 0.37 Å; Tritschler et al., 2009), indicating that the domains do not undergo major structural rearrangements upon binding.

Most importantly, the arrangement of the CNOT1 MIF4G and DDX6 RecA-C domains is highly similar to the arrangement of the complex of the eIF4G MIF4G domain bound to eIF4A RecA-C (Figures 5E and 5F; Schütz et al., 2008), adding to a growing number of structurally similar, yet specific, MIF4G-RecA-C complexes (Buchwald et al., 2013). Similar to the eIF4G-eIF4A interaction, the RecA-C domain binds the concave surface of the MIF4G domain, and the interface is formed by equivalent secondary structural elements. In particular, helix α1 and the loops L3, L5, and L7 of the RecA-C domains interact with the first two N-terminal HEAT repeats of the MIF4G domain, including mainly residues from helix α1 and α3 plus their adjacent interrepeat loops L2 and L4 (Figures 5G and 5H).

The interface area in the CNOT1-DDX6 complex is relatively small (640 $Å^2$) and dominated by polar interactions. Of particular interest is R375 in loop L7 of DDX6 (Figure 5G), which inserts deeply into a highly conserved patch in CNOT1 and makes hydrogen bonds with the main-chain oxygens of CNOT1 residues F1103 (helix α1) and L1106 (loop L2) while stacking onto F1145 (loop L4). This arginine is remarkably conserved in eIF4A1, eIF4A2, eIF4A3, and other DEAD-box proteins (Figure S6I) and plays an equivalent role in mediating interactions with MIF4G domains (Buchwald et al., 2013; Schütz et al., 2008). Consequently, this "arginine anchor" likely contributes significantly to the affinity of MIF4G-RecA-C interactions, but does not explain the specificity of individual pairs. This situation is different for residues R320, Q322, and N324 in loop L3 of DDX6 (Figures 5H and 5I). Although the length of this specificity loop is identical in eIF4A1-3 and other DEAD-box helicases, the sequence is not, thereby allowing the formation of unique polar contacts to residues E1097, R1138, E1142, and N1144 of CNOT1 and explaining the preference of CNOT1 for DDX6 (Figures 5H, 5I, and S6I).

To validate the interface experimentally, we mutated specific interface residues to alanine and performed pull-downs and coimmunoprecipitation assays. In DDX6, we substituted the anchor arginine (R375) or both Q322 and N324 (Q,N) from the specificity loop. These mutations abolished or strongly reduced binding of DDX6 to CNOT1 full length, the CNOT1-M fragment, or the MIF4G domain (Figures 6A, 6B, and S6E–S6G). In CNOT1, the deletion of helix α1 and loop L2 (CNOT1 ΔhI) led

to a reduced interaction with DDX6, which was disrupted completely when we additionally substituted R1138 (in α3) and N1144 and F1145 in loop L4 with alanine (Figure 6C, lanes 9–12, and S6H). Importantly, the mutations did not destroy the fold of the MIF4G domain because CAF1 binding was not affected (Figure S6H).

A Role for the CNOT1-DDX6 Interface in Silencing

To test the functional relevance of the CNOT1-DDX6 interaction in silencing, we used the complementation assay described above with the exception that DDX6 was depleted. Western blot analysis indicated that the levels of DDX6 in the depleted cells were reduced to 10% of the control levels (Figure 6D), without affecting endogenous AGO2 and TNRC6A expression (Figure S5N). In DDX6-depleted cells, silencing of the psiCHECK-8×Let-7 reporter was suppressed, leading to a 7-fold increase in F-Luc activity (Figure 6E), in agreement with previous studies (Chu and Rana, 2006; Eulalio et al., 2007). Silencing was rescued by expression of an shRNA-resistant version of wild-type DDX6, but not by the DDX6 R375A mutant, whereas the Q322A N324A double mutant was partially active (Figure 6E), in agreement with its residual binding to CNOT1 (Figure S6E, lane 12). The DDX6 proteins were expressed at similar levels (Figure 6F). These levels were comparable to the levels of endogenous DDX6 in control cells (Figure 6D, lane 6 versus lane 1). Our data support a role for the DDX6-CNOT1 interaction in silencing.

The CNOT1 MIF4G Domain Is a Central Node for CCR4-NOT Function

Structural superposition of the human CNOT1 MIF4G-DDX6 and CNOT1 MIF4G-CAF1 complexes demonstrates that CNOT1 can bind DDX6 and the CAF1 deadenylase simultaneously (Figures 5J and S6J). Similar superpositions show that the interaction of the DDX6 RecA-C domain with the isolated CNOT1 MIF4G domain does not interfere structurally with the recruitment of either EDC3, LSm14A, or Pat to their common binding surface on DDX6 RecA-C (reviewed by Jonas and Izaurralde, 2013). Together, these interactions provide a missing direct physical link between the major deadenylation complex (CCR4-NOT) and the decapping network where EDC3 and Pat act as decapping activators. In other words, the interaction of DDX6 with the CNOT1 MIF4G domain provides a plausible molecular explanation for the coupling of deadenylation to decapping and for the ability of the CCR4-NOT complex to repress translation in the absence of deadenylation. TNRC6-binding to the adjacent CN9BD-CNOT9 complex illustrates how the CCR4-NOT complex is recruited to miRNA targets.

DISCUSSION

The mechanisms by which the CCR4-NOT complex is recruited to miRNA targets and represses translation have remained elusive. Here, we show that the middle region of CNOT1 assists in the recruitment of the complex to miRNA targets via W-binding pockets in the CN9BD-CNOT9 module and orchestrates deadenylation, translational repression, and decapping via the MIF4G domain, which can bind to the CAF1 and CCR4 deadenylases and to the translational repressor and decapping activator DDX6. Notably, similar findings are reported by Mathys et al. (2014) in this issue.

The Functional Repertoire of CNOT9

In this study, we reveal an unexpected role for the highly conserved CNOT9 subunit of the CCR4-NOT complex in miRNA-mediated gene silencing. CNOT9 is required for retinoic acid-induced cell differentiation in mammals and is overexpressed in breast cancer cells (Hiroi et al., 2002; Ajiro et al., 2009). However, it remains unclear whether the role of CNOT9 in cell differentiation and proliferation is linked to or independent of its role as subunit of the CCR4-NOT complex. Through our structural analysis, we have generated CNOT9 mutants that retain the ability to homodimerize but are not incorporated in the CCR4-NOT complex, which provides an important tool to study CNOT9 function in vivo.

CNOT9 interacts with several protein partners in different cellular contexts. These interacting partners include Grb10 interacting proteins GIGYF1 and GIGYF2 that are involved in epidermal growth factor receptor (EGFR) signaling (Ajiro et al., 2009). Notably, GIGYF2 (also known as TNRC15) is a W-rich protein, which raises the possibility that some CNOT9 partners may compete with TNRC6s for binding to the W pockets. Clearly, the identification of the W-binding pockets increases the functional repertoire of CNOT9 that needs to be considered when studying CNOT9 function.

Interaction of GW182 Proteins with Their Binding Partners

GW182 proteins interact with AGOs and the PAN2-PAN3 and CCR4-NOT deadenylase complexes through W-containing motifs (Fabian and Sonenberg, 2012; Braun et al., 2013). Previous studies on the interaction of these motifs with AGOs and PAN3 indicate that molecular recognition is predominantly restricted to the W residues, which insert into hydrophobic pockets on the respective protein partners (Schirle and MacRae, 2012; Christie et al., 2013; Pfaff et al., 2013).

Here, we show that this mode of molecular recognition is also observed for the interaction of GW182 proteins with the CN9BD-CNOT9 complex. This raises the question of how binding specificity and affinity are achieved. One possibility is that the flanking sequences and the spatial arrangement of the Ws contribute to the affinity of the interaction (Schirle and MacRae, 2012; Pfaff et al., 2013). For example, a common feature of the W-binding pockets in AGO2 and CNOT9 is that the spatial arrangement of the pockets is similar, and both proteins can accommodate consecutive W residues provided that they are at least 8–10 residues apart (this study; Schirle and MacRae, 2012; Pfaff et al., 2013). Concerted binding of adjacent W residues may contribute to the affinity of the interaction via additive or avidity effects.

A remarkable aspect of the interaction of GW182 proteins with the CCR4-NOT complex is that it involves multiple binding sites (in addition to the ones identified on CNOT9), presumably leading to higher affinity and specificity. The precise location of these additional sites remains unclear because the isolated CNOT1-N, MIF4G, and CNOT1-C regions do not detectably interact with TNRC6s in coimmunoprecipitation assays. However, these

regions seem to synergize in the context of full-length CNOT1 because a CNOT1 mutant that does not bind CNOT9 retains full TNRC6-binding capacity (Figures 4E and S5M). A key direction for future work will be to identify these multiple and redundant binding sites and determine how they contribute to the recruitment of the CCR4-NOT complex to miRNA targets.

CNOT1-M Coordinates Deadenylation, Translational Repression, and Decapping

MIF4G domains are present in a wide variety of proteins and have been shown to interact with and regulate the activity of DEAD-box proteins (Buchwald et al., 2013). The MIF4G domain of eIF4G specifically binds to eIF4A1 (Schütz et al., 2008). Remarkably, other MIF4G domain-containing proteins, such as DAP5, CWC22, and Gle1, interact with the RecA-C domains of eIF4A1, eIF4A3, and Dbp5, respectively, using a recognition mode similar to that observed in the eIF4G-eIF4A complex (reviewed by Buchwald et al., 2013). Based on these observations, it has been proposed that the CNOT1 MIF4G domain interacts with eIF4A2 (Meijer et al., 2013). Here, we show that the CNOT1 MIF4G domain interacts preferentially with DDX6. The binding specificity is imparted by a few amino acid substitutions, in particular in loop L3 of the RecA-C domains, which establish specific hydrogen bonds (Figure 5I).

DDX6 plays a role in repressing translation by slowing translation elongation (Presnyak and Coller, 2013). Its role as an activator of decapping has been proposed to be an indirect consequence of the inhibition of translation (Presnyak and Coller, 2013). However, DDX6 interacts directly with the decapping factors EDC3 and Pat and indirectly with the catalytic core of the decapping complex formed by the decapping enzyme DCP2 and its coactivator DCP1 (Jonas and Izaurralde, 2013). Thus, DDX6 could play a direct role in decapping by promoting recruitment of decapping complexes to the mRNA target. These observations, together with our studies, indicate that the CNOT1-DDX6 complex provides a missing direct physical link between deadenylation and decapping.

The structure of the NOT1-DDX6 complex together with available structures of DDX6 bound to EDC3, EDC3 bound to metazoan DCP1, and the Dcp1-Dcp2 complex (reviewed by Jonas and Izaurralde, 2013) present snapshots of consecutive steps in the 5′-to-3′ mRNA decay pathway. Along with the structure of the CN9BD-CNOT9 complex, these structures establish a chain of physical interactions to describe in molecular terms how the CCR4-NOT complex is recruited to miRNA targets and enrolls a translational repressor, which in turn engages the decapping machinery (Figures 5J and S6J).

EXPERIMENTAL PROCEDURES

DNA constructs are described in detail in the Supplemental Information. Mutations used in this study are listed in Table S1.

Pulldowns, Coimmunoprecipitation Assays, and Western Blotting

Coimmunoprecipitation assays in human and *Dm* S2 cells were performed as previously described (Bawankar et al., 2013; Braun et al., 2011). Antibodies used in this study are listed in Table S1. All western blots were developed using the ECL Western Blotting Detection System (GE Healthcare). A detailed protocol for the pull-down assays using recombinant proteins can be found in the Supplemental Experimental Procedures.

Protein Expression and Purification

The CNOT1 CN9BD and the CNOT9 ARM domain were coexpressed in *E. coli* BL21 Star (DE3) cells (Invitrogen) and induced at an optical density 600 (OD_{600}) of 0.6 with 1 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG). Following induction, the proteins were expressed overnight at 20°C. The complex was purified using amylose resin, a heparin column, and finally by size-exclusion chromatography as described in the Supplemental Experimental Procedures. The CNOT1 MIF4G domain and the DDX6 RecA-C domain were expressed separately in *E. coli* Rosetta 2 cells (Novagen). The proteins were purified as described previously (Petit et al., 2012; Tritschler et al., 2009), mixed at a ratio of 1:1.2 (CNOT1:DDX6), and concentrated to 10 mg/ml in 10 mM HEPES (pH 7.0), 300 mM NaCl, and 5 mM dithiothreitol (DTT).

Crystallization, Data Collection, and Structure Determination

Crystals of the CNOT1 CN9BD-CNOT9 ARM complex were obtained using the hanging-drop and sitting-drop vapor diffusion methods over a 500 μl reservoir at 18°C. The protein solution was mixed in a 1:1 ratio (0.8 μl + 0.8 μl) with a reservoir solution containing 100 mM MES (pH 6.0), 8% polyethylene glycol (PEG) 6,000, and 80 mM $MgCl_2$. Alternatively, crystals were obtained by mixing the protein solution in a 1:1 ratio (0.8 μl + 0.8 μl) with the reservoir solution containing 100 mM MES (pH 6.0), 11% PEG 6,000, and 50 mM $MgCl_2$ supplemented with 40 mM L-tryptophan. Crystals of the CNOT1 MIF4G-DDX6 RecA-C complex were obtained using the hanging-drop vapor diffusion method over a 500 μl reservoir (100 mM Tris (pH 8.0) and 16% PEG 6,000) at 20°C by mixing 1.5 μl of the protein solution with 1.5 μl of the reservoir solution. All crystals were cryoprotected using the corresponding reservoir solution supplemented with 25% glycerol and subsequently flash frozen in liquid nitrogen.

All diffraction data sets were recorded on a PILATUS 6 M detector at the PXII beamline of the Swiss Light Source (SLS) at a temperature of 100 K. A detailed description of the structure determination process can be found in the Supplemental Experimental Procedures. The refinement statistics are summarized in Table 1.

Complementation Assays in Human Cells

Knockdowns and complementation assays were performed as described in the Supplemental Experimental Procedures.

ACCESSION NUMBERS

The atomic coordinates of CNOT1-CN9BD bound to the CNOT9 ARM repeat in the absence or presence of tryptophan and of CNOT1 MIF4G bound to the DDX6 RecA-C domain were deposited in the Protein Data Bank (PDB) under ID codes 4CRU, 4CRV, and 4CRW, respectively.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, six figures, and one table and can be found with this article online at http://dx.doi.org/10.1016/j.molcel.2014.03.034.

AUTHOR CONTRIBUTIONS

Y.C., A.B., and D.K.-Ö. contributed equally to this work. Y.C. and A.B. purified, crystallized, and solved the structures of CNOT1-CN9BD bound to the CNOT9 ARM repeat and of the CNOT1 MIF4G bound to DDX6 RecA-C domain. Y.C., A.B., and O.W. collected and analyzed diffraction data. Y.C. and A.B. performed pull-downs in vitro. D.K.-Ö. performed coimmunoprecipitations and functional assays in human cells. P.B. performed coimmunoprecipitation assays in S2 cells. B.L. performed the DDX6 complementation assays. C.-T.C. performed DDX6 coimmunoprecipitation assays in human cells. E.I. conceived and supervised the project. A.B., E.I., and O.W. wrote the manuscript. All authors corrected the manuscript.

ACKNOWLEDGMENTS

We thank M. Fauser and S. Helms for excellent technical assistance. We are grateful to M. Christie, H. Budde, E. Huntzinger, and L. Wohlbold for generating some of the constructs used in this study; E. Huntzinger for performing initial experiments; R. Büttner for the setup of crystallization screens; Y. Tomari for providing the psiCHECK-8×Let-7 miRNA reporter; and the staff at the PX beamlines of the Swiss Light Source for assistance with data collection. This work was supported by the Max Planck Society and by grants from the Deutsche Forschungsgemeinschaft (DFG, FOR855, and the Gottfried Wilhelm Leibniz Program awarded to E.I.).

Received: January 24, 2014
Revised: February 13, 2014
Accepted: March 18, 2014
Published: April 24, 2014

REFERENCES

Ajiro, M., Katagiri, T., Ueda, K., Nakagawa, H., Fukukawa, C., Lin, M.L., Park, J.H., Nishidate, T., Daigo, Y., and Nakamura, Y. (2009). Involvement of RQCD1 overexpression, a novel cancer-testis antigen, in the Akt pathway in breast cancer cells. Int. J. Oncol. *35*, 673–681.

Ameres, S.L., and Zamore, P.D. (2013). Diversifying microRNA sequence and function. Nat. Rev. Mol. Cell Biol. *14*, 475–488.

Bawankar, P., Loh, B., Wohlbold, L., Schmidt, S., and Izaurralde, E. (2013). NOT10 and C2orf29/NOT11 form a conserved module of the CCR4-NOT complex that docks onto the NOT1 N-terminal domain. RNA Biol. *10*, 228–244.

Braun, J.E., Huntzinger, E., Fauser, M., and Izaurralde, E. (2011). GW182 proteins directly recruit cytoplasmic deadenylase complexes to miRNA targets. Mol. Cell *44*, 120–133.

Braun, J.E., Huntzinger, E., and Izaurralde, E. (2013). The role of GW182 proteins in miRNA-mediated gene silencing. Adv. Exp. Med. Biol. *768*, 147–163.

Buchwald, G., Schüssler, S., Basquin, C., Le Hir, H., and Conti, E. (2013). Crystal structure of the human eIF4AIII-CWC22 complex shows how a DEAD-box protein is inhibited by a MIF4G domain. Proc. Natl. Acad. Sci. USA *110*, E4611–E4618.

Chekulaeva, M., Mathys, H., Zipprich, J.T., Attig, J., Colic, M., Parker, R., and Filipowicz, W. (2011). miRNA repression involves GW182-mediated recruitment of CCR4-NOT through conserved W-containing motifs. Nat. Struct. Mol. Biol. *18*, 1218–1226.

Christie, M., Boland, A., Huntzinger, E., Weichenrieder, O., and Izaurralde, E. (2013). Structure of the PAN3 pseudokinase reveals the basis for interactions with the PAN2 deadenylase and the GW182 proteins. Mol. Cell *51*, 360–373.

Chu, C.Y., and Rana, T.M. (2006). Translation repression in human cells by microRNA-induced gene silencing requires RCK/p54. PLoS Biol. *4*, e210.

Coller, J.M., Tucker, M., Sheth, U., Valencia-Sanchez, M.A., and Parker, R. (2001). The DEAD box helicase, Dhh1p, functions in mRNA decapping and interacts with both the decapping and deadenylase complexes. RNA *7*, 1717–1727.

Eulalio, A., Rehwinkel, J., Stricker, M., Huntzinger, E., Yang, S.F., Doerks, T., Dorner, S., Bork, P., Boutros, M., and Izaurralde, E. (2007). Target-specific requirements for enhancers of decapping in miRNA-mediated gene silencing. Genes Dev. *21*, 2558–2570.

Fabian, M.R., and Sonenberg, N. (2012). The mechanics of miRNA-mediated gene silencing: a look under the hood of miRISC. Nat. Struct. Mol. Biol. *19*, 586–593.

Fabian, M.R., Cieplak, M.K., Frank, F., Morita, M., Green, J., Srikumar, T., Nagar, B., Yamamoto, T., Raught, B., Duchaine, T.F., and Sonenberg, N. (2011). miRNA-mediated deadenylation is orchestrated by GW182 through two conserved motifs that interact with CCR4-NOT. Nat. Struct. Mol. Biol. *18*, 1211–1217.

Garces, R.G., Gillon, W., and Pai, E.F. (2007). Atomic model of human Rcd-1 reveals an armadillo-like-repeat protein with in vitro nucleic acid binding properties. Protein Sci. *16*, 176–188.

Hata, H., Mitsui, H., Liu, H., Bai, Y.L., Denis, C.L., Shimizu, Y., and Sakai, A. (1998). Dhh1p, a putative RNA helicase, associates with the general transcription factors Pop2p and Ccr4p from Saccharomyces cerevisiae. Genetics *148*, 571–579.

Hiroi, N., Ito, T., Yamamoto, H., Ochiya, T., Jinno, S., and Okayama, H. (2002). Mammalian Rcd1 is a novel transcriptional cofactor that mediates retinoic acid-induced cell differentiation. EMBO J. *21*, 5235–5244.

Huntzinger, E., and Izaurralde, E. (2011). Gene silencing by microRNAs: contributions of translational repression and mRNA decay. Nat. Rev. Genet. *12*, 99–110.

Huntzinger, E., Kuzuoglu-Öztürk, D., Braun, J.E., Eulalio, A., Wohlbold, L., and Izaurralde, E. (2013). The interactions of GW182 proteins with PABP and deadenylases are required for both translational repression and degradation of miRNA targets. Nucleic Acids Res. *41*, 978–994.

Iwasaki, S., Kawamata, T., and Tomari, Y. (2009). Drosophila argonaute1 and argonaute2 employ distinct mechanisms for translational repression. Mol. Cell *34*, 58–67.

Jonas, S., and Izaurralde, E. (2013). The role of disordered protein regions in the assembly of decapping complexes and RNP granules. Genes Dev. *27*, 2628–2641.

Kawashima, S., Pokarowski, P., Pokarowska, M., Kolinski, A., Katayama, T., and Kanehisa, M. (2008). AAindex: amino acid index database, progress report 2008. Nucleic Acids Res. *36* (Database issue), D202–D205.

Maillet, L., and Collart, M.A. (2002). Interaction between Not1p, a component of the Ccr4-not complex, a global regulator of transcription, and Dhh1p, a putative RNA helicase. J. Biol. Chem. *277*, 2835–2842.

Mathys, H., Basquin, J., Ozgur, S., Czarnocki-Cieciura, M., Bonneau, F., Aartse, A., Dziembowski, A., Nowotny, M., Conti, E., and Filipowicz, W. (2014). Structural and Biochemical Insights to the Role of the CCR4-NOT Complex and DDX6 ATPase in MicroRNA Repression. Mol. Cell *54*. Published online April 24, 2014. http://dx.doi.org/10.1016/j.molcel.2014.03.036.

Meijer, H.A., Kong, Y.W., Lu, W.T., Wilczynska, A., Spriggs, R.V., Robinson, S.W., Godfrey, J.D., Willis, A.E., and Bushell, M. (2013). Translational repression and eIF4A2 activity are critical for microRNA-mediated gene regulation. Science *340*, 82–85.

Nishihara, T., Zekri, L., Braun, J.E., and Izaurralde, E. (2013). miRISC recruits decapping factors to miRNA targets to enhance their degradation. Nucleic Acids Res. *41*, 8692–8705.

Petit, A.P., Wohlbold, L., Bawankar, P., Huntzinger, E., Schmidt, S., Izaurralde, E., and Weichenrieder, O. (2012). The structural basis for the interaction between the CAF1 nuclease and the NOT1 scaffold of the human CCR4-NOT deadenylase complex. Nucleic Acids Res. *40*, 11058–11072.

Pfaff, J., Hennig, J., Herzog, F., Aebersold, R., Sattler, M., Niessing, D., and Meister, G. (2013). Structural features of Argonaute-GW182 protein interactions. Proc. Natl. Acad. Sci. USA *110*, E3770–E3779.

Presnyak, V., and Coller, J. (2013). The DHH1/RCKp54 family of helicases: an ancient family of proteins that promote translational silencing. Biochim. Biophys. Acta *1829*, 817–823.

Schirle, N.T., and MacRae, I.J. (2012). The crystal structure of human Argonaute2. Science *336*, 1037–1040.

Schütz, P., Bumann, M., Oberholzer, A.E., Bieniossek, C., Trachsel, H., Altmann, M., and Baumann, U. (2008). Crystal structure of the yeast eIF4A-eIF4G complex: an RNA-helicase controlled by protein-protein interactions. Proc. Natl. Acad. Sci. USA *105*, 9564–9569.

Temme, C., Zhang, L.B., Kremmer, E., Ihling, C., Chartier, A., Sinz, A., Simonelig, M., and Wahle, E. (2010). Subunits of the Drosophila CCR4-NOT complex and their roles in mRNA deadenylation. RNA *16*, 1356–1370.

Tritschler, F., Braun, J.E., Eulalio, A., Truffault, V., Izaurralde, E., and Weichenrieder, O. (2009). Structural basis for the mutually exclusive anchoring of P body components EDC3 and Tral to the DEAD box protein DDX6/Me31B. Mol. Cell *33*, 661–668.

Wahle, E., and Winkler, G.S. (2013). RNA decay machines: deadenylation by the Ccr4-not and Pan2-Pan3 complexes. Biochim. Biophys. Acta *1829*, 561–570.

Yoder-Hill, J., Pause, A., Sonenberg, N., and Merrick, W.C. (1993). The p46 subunit of eukaryotic initiation factor (eIF)-4F exchanges with eIF-4A. J. Biol. Chem. *268*, 5566–5573.

Zekri, L., Kuzuoğlu-Öztürk, D., and Izaurralde, E. (2013). GW182 proteins cause PABP dissociation from silenced miRNA targets in the absence of deadenylation. EMBO J. *32*, 1052–1065.

Molecular Cell, Volume *54*

Supplemental Information

A DDX6-CNOT1 Complex and W-Binding Pockets in CNOT9 Reveal Direct Links between miRNA Target Recognition and Silencing

Ying Chen, Andreas Boland, Duygu Kuzuoğlu-Öztürk, Praveen Bawankar, Belinda Loh, Chung-Te Chang, Oliver Weichenrieder, and Elisa Izaurralde

SUPPLEMENTAL FIGURES

Figure S1

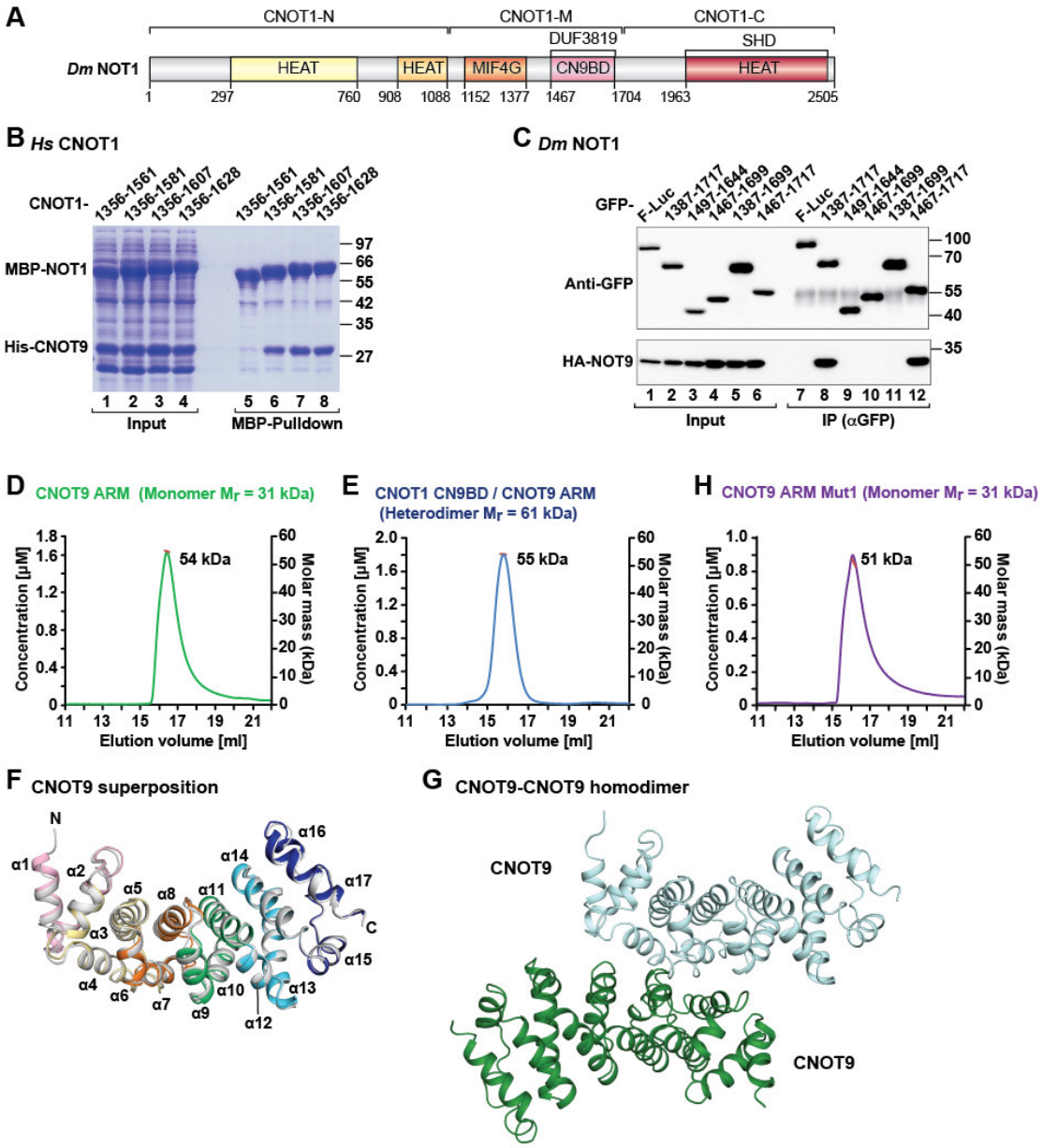


Figure S1, related to Figure 1. Interaction of NOT1 with NOT9 and MALLS analysis

(A) Domain organization of *Dm* NOT1. Abbreviations are as in Figure 1A.

(B) MBP pulldown showing the direct interaction between recombinant MBP-CNOT1 fragments and the His-tagged CNOT9 ARM-repeat domain (residues 19–285). See also Figure 1B,C.

(C) Western blot showing the interaction between GFP-NOT1 fragments and HA-

tagged NOT9 ARM-repeat domain in *Dm* S2 cells.

(D,E,H) Analytical size exclusion chromatography and multi-angle laser light scattering (MALLS). Isolated CNOT9 (D) or its CNOT9 Mut1 (H) form homodimers, which are incompatible with the CN9BD-CNOT9 binary complex (E). The expected molecular weights of the proteins (M_r) and the values measured in solution are indicated.

(F) Superposition of the CNOT9 ARM domain from the CNOT1-CNOT9 complex with CNOT1 (colored as in Figure 1F) with isolated CNOT9 (gray; PDB code 2FV2). See also Figure 1D–G.

(G) Cartoon representation of the CNOT9 ARM-repeat homodimer (PDB code 2FV2; Garces et al., 2007). The CNOT9 molecule in cyan is shown in the same orientation as in Figure 1D.

Figure S2

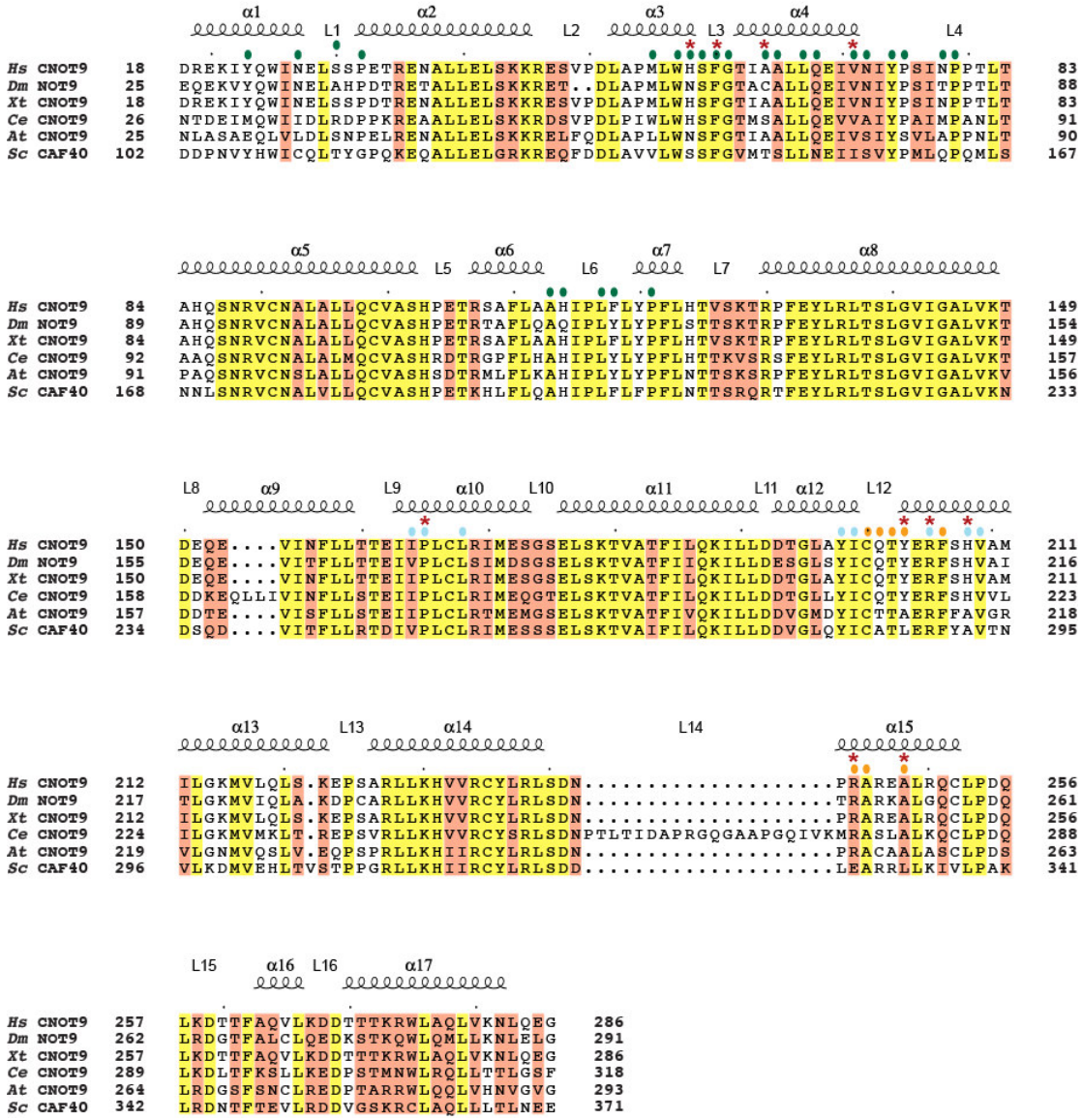


Figure S2, related to Figure 2. Structure-based multiple sequence alignment of the CNOT9 ARM-repeat domain

Secondary structure elements as determined from the structure are shown above the alignment. Residues conserved in all aligned sequences are shown with a yellow background, and residues with >70% similarity are highlighted in orange. Residues that form the interface with CNOT1 are indicated by green dots. Residues in pocket 1 and 2 are indicated by cyan and orange dots, respectively. Residues mutated in this study are marked by red asterisks. The species abbreviations are as follows: *Hs* (*Homo sapiens*), *Dm* (*Drosophila melanogaster*), *Xt* (*Xenopus tropicalis*), *Ce*

(*Caenorhabditis elegans*), *At* (*Arabodopsis thaliana*) and *Sc* (*Saccharomyces cerevisiae*). See also Figures 1–3. The alignments were carried out using ESPript (Gouet et al., 2003).

Figure S3

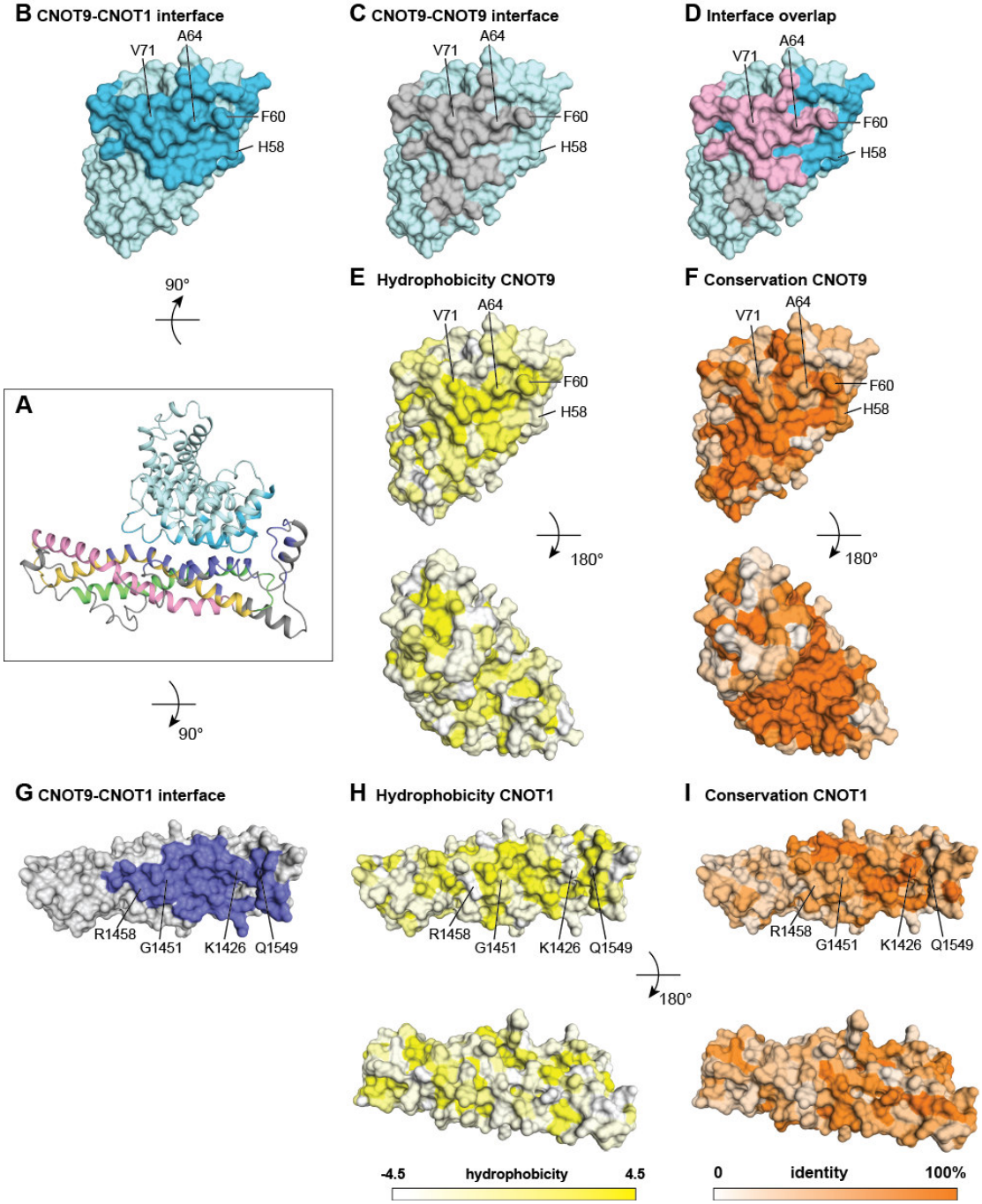


Figure S3, related to Figures 2 and 3. Conservation and hydrophobicity of the CN9BD-CNOT9 interface

(A) Cartoon representation of the CN9BD-CNOT9 complex (oriebtation as in Figure 1E). To visualize the interacting surfaces, the binding partners seen in (A) were rotated around the horizontal axis by 90º; upwards for the surface of CNOT9 (B–F), and downwards for the surface of the CN9BD (G–I). Selected interface residues are labeled for orientation.

(B–D) View of the CNOT9 surface that binds either CNOT9 in the homodimer or CNOT1 in the binary complex. The residues involved in CNOT1 binding and CNOT9 homodimerization are shown in dark cyan (B) and gray (C), respectively. Panel D shows a superposition of the views in panels B and C. The overlap between the two binding interfaces is shown in salmon. See also Figure 2.

(E and F) Surface representation of CNOT9 colored according to hydrophobicity (E) and conservation (F). The upper panels show a view of the CNOT1-binding interface in the same orientation as in panel (B). The lower panels show the opposite surface.

(G) View of the CN9BD surface that binds CNOT9 in the binary complex. The residues involved in CNOT9 binding are shown in purple.

(H and I) Surface representation of the CN9BD colored according to hydrophobicity (H) and conservation (I). The upper panels show a view in the same orientation as in panel (G). The lower panels show the opposite surface. The conservation scores of the individual residues are represented on the surface by color gradients from light (no conservation) to dark orange. Surfaces are colored white to yellow with increasing hydrophobicity as described in Figure 3.

Figure S4

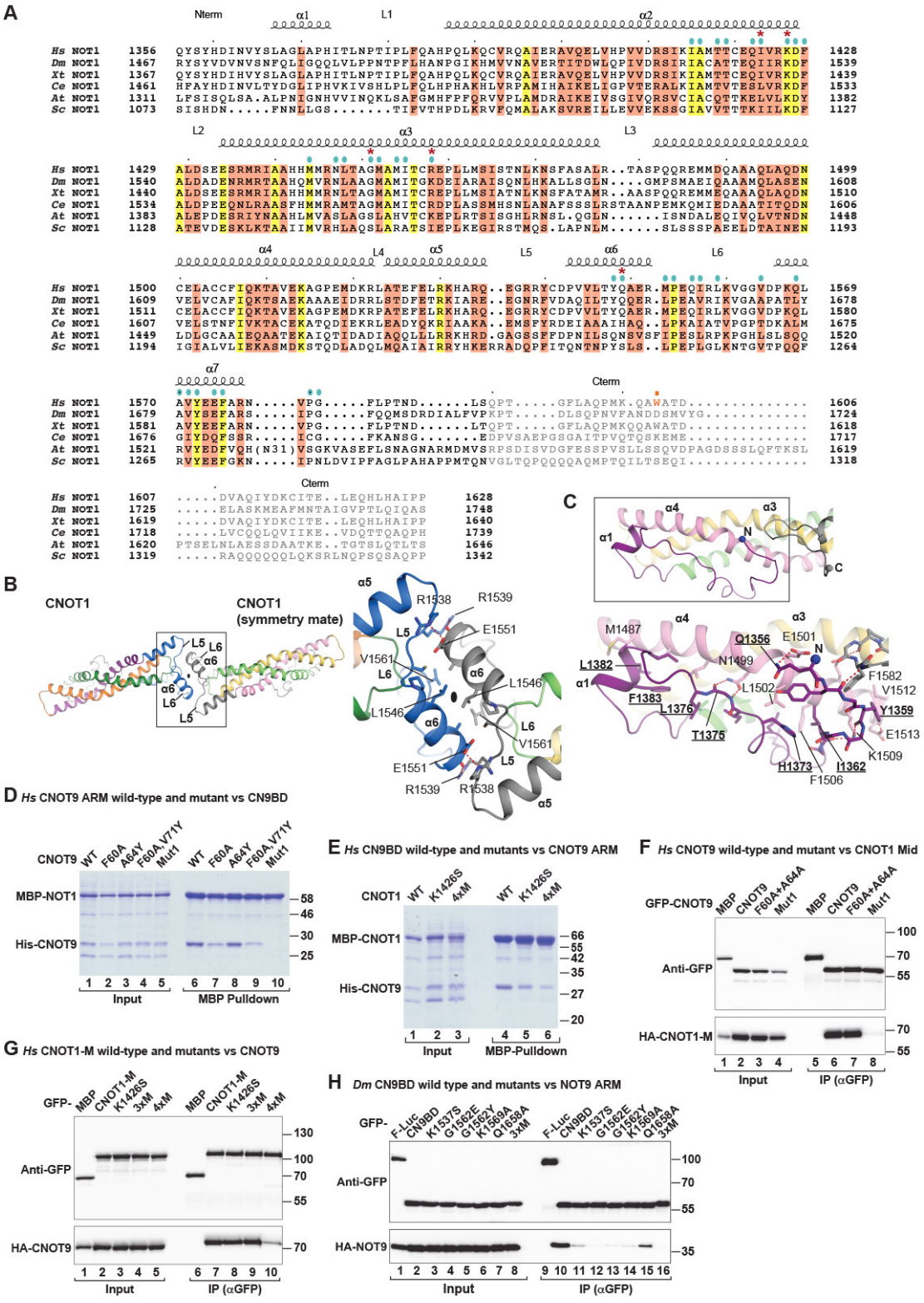


Figure S4, related to Figure 4. Structure-based sequence alignment of the CNOT1 CN9BD and interaction with CNOT9

(A) Secondary structure elements as determined from the CN9BD structure are shown above the alignment. Residues interacting with CNOT9 are indicated by cyan dots.

Residues that are not visible in the structure are shown in gray. Residue W1603, which inserts into the W-binding pocket in CNOT9, is shown in orange. Colors and species abbreviations are as described in Figure S2. See also Figures 1–4.

(B) Cartoon representation showing the crystal packing (left panel) and close-up view on the interactions between symmetry mates in the crystal lattice around a two-fold axis (right panel).

(C) Cartoon representation of the CN9BD (top panel) and close-up view (lower panel) showing how the N-terminal extension folds back and interacts with helix α4. N-terminal residues are shown in bold and are underlined. See also Figure 2.

(D and E) MBP pulldown using recombinant MBP-CNOT1 CN9BD (wild-type or the indicated mutants) and His-tagged CNOT9 ARM-repeat domain (wild-type or the indicated mutants).

(F) Western blot showing the interaction between GFP-CNOT9 (wild-type or the indicated mutants) and HA-CNOT1-M region in HEK293T cells. See also Figure S2.

(G) Western blot showing the interaction between GFP-CNOT1-M (wild-type or the indicated mutants) and HA-MBP-CNOT9 in HEK293T cells.

(H) Interaction between GFP-NOT1 CN9BD (wild-type or the indicated mutants) and HA-NOT9-ARM repeat domain in *Dm* S2 cells.

Figure S5

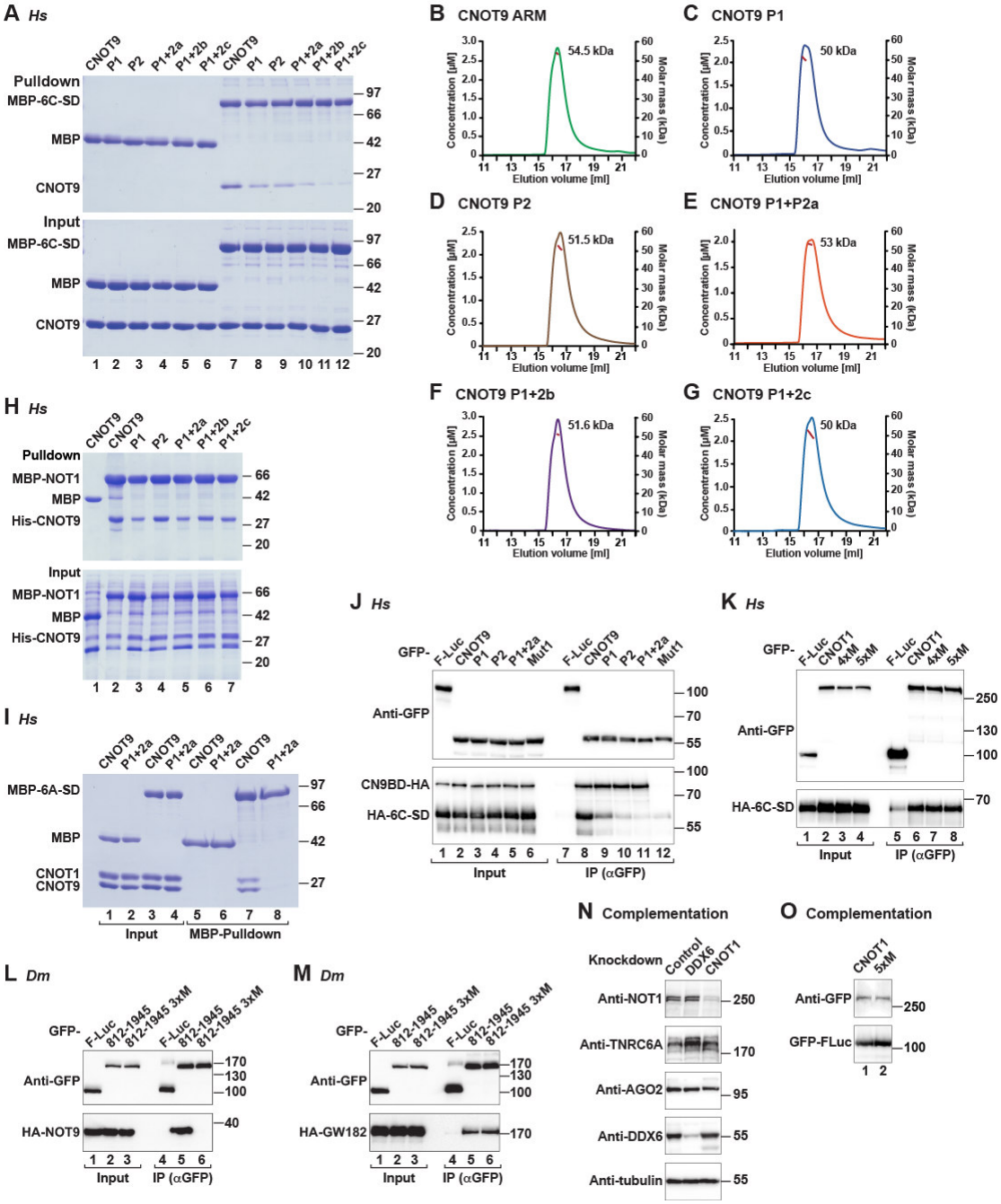


Figure S5, related to Figures 4 and 6. NOT9 W-binding pockets interact with GW182/TNRC6 proteins

(A) MBP pulldowns using recombinant MBP-tagged 6C-SD and His-tagged CNOT9 ARM-repeat domain (wild-type or the indicated pocket mutants). MBP served as a negative control. See also Figure 4 and Table S1.

(B-G) Analytical size exclusion chromatography and MALLS analysis of the wild-type CNOT9 homodimers or the indicated pocket mutants. The molecular weight of

the proteins measured in solution is indicated. The expected molecular weight for a dimer is 62 kDa.

(H) MBP pulldown showing the interaction of MBP-CNOT1 CN9BD with His-tagged CNOT9 ARM-repeat domain (wild-type or the indicated W-binding pocket mutants).

(I) MBP pulldown using recombinant MBP-tagged 6A-SD and pre-assembled CNOT1 CN9BD-CNOT9 ARM complexes containing CNOT9 wild-type or the P1+2a mutant. MBP served as a negative control.

(J) Interaction between GFP-CNOT9 (wild-type or the indicated mutants) and HA-6C-SD in the presence of HA-MBP tagged CN9BD in HEK293T cells.

(K) Interaction between GFP-CNOT1 (wild-type or the 4xM and 5xM mutants that do not bind CNOT9) and HA-6C-SD in HEK293T cells. The mutations are described in Table S1.

(L and M) Interaction of GFP-*Dm* NOT1 (residues 812-1945, either wild-type or the 3xM mutant) with CNOT9 (L) and *Dm* GW182 (M).

(N) Western blot analysis showing that the expression of endogenous AGO2 and TNRC6A is not affected in cells depleted of either CNOT1 or DDX6. The samples correspond to the depletions shown in Figures 4G and 6E.

(O) Western blot analysis showing the equivalent expression of the CNOT1 proteins used in the complementation assay shown in Figure 4G.

Figure S6

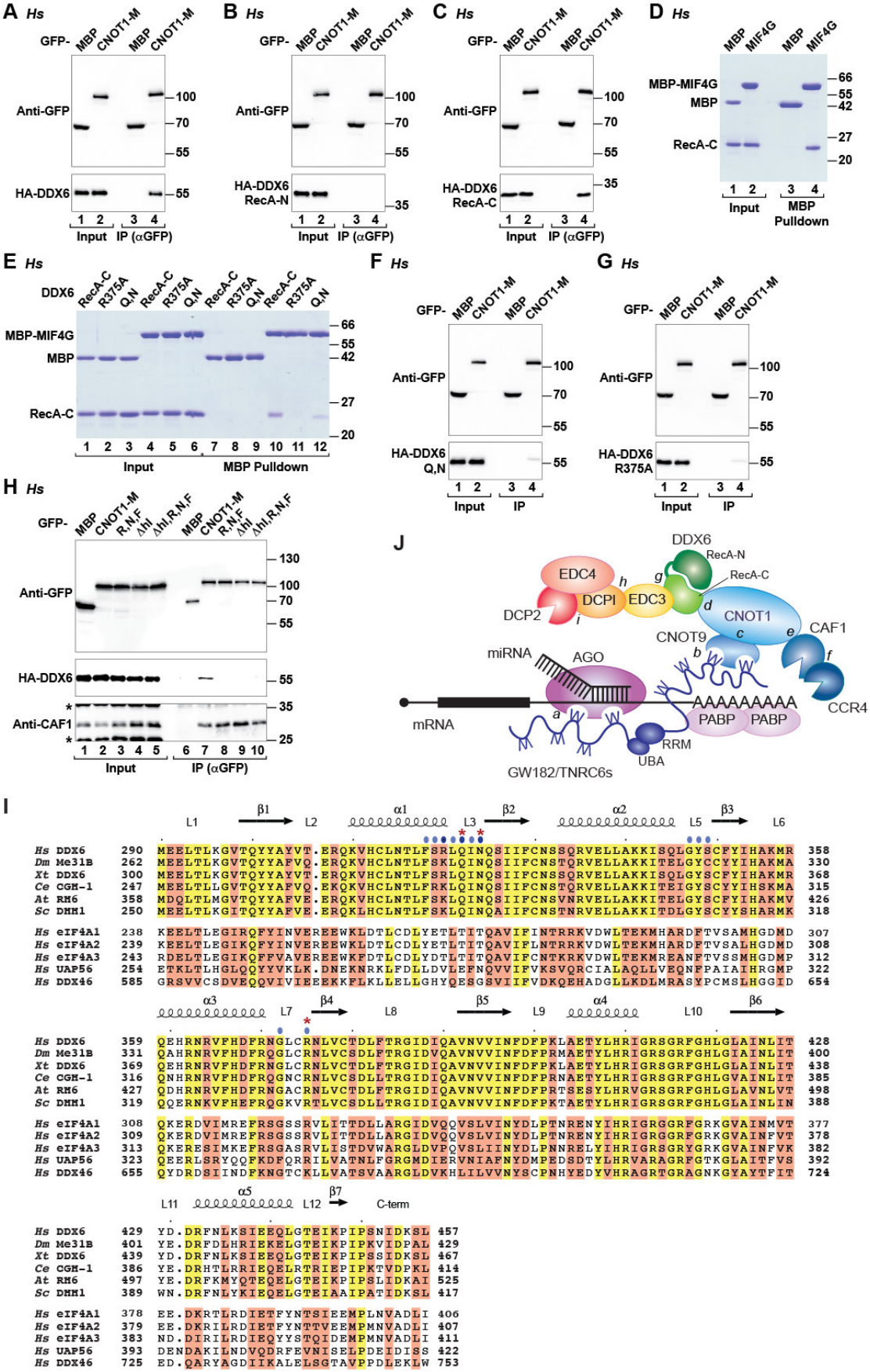


Figure S6, related to Figures 5 and 6. The CNOT1 MIF4G-DDX6 RecA-C

interaction

(A-C) Interaction between GFP-CNOT1-M and either full-length HA-tagged DDX6 (A), the RecA-N (B) or RecA-C (C) domains in HEK293T cells. See also Figure 5.

(D and E) MBP pulldowns using recombinant MBP-tagged CNOT1 MIF4G domain and His-tagged DDX6 RecA-C domain (wild-type or mutants). MBP served as a negative control.

(F and G) Interaction between GFP-CNOT1-M and HA-tagged wild-type DDX6 or the indicated DDX6 mutants in HEK293T cells. The coimmunoprecipitation shown in panel (A) was performed in parallel and served as positive control.

(H) Interaction between GFP-CNOT1-M (wild-type or the indicated mutants) and HA-tagged DDX6 in HEK293T cells. The presence of endogenous CAF1 in the immunoprecipitates was confirmed using specific anti-CAF1 antibodies. The asterisk indicates crossreactivity of the antibody with an endogenous protein in the input lysate. See also Figures 5 and 6, Figure S8 and Table S1.

(I) Structure-based multiple sequence alignment of DDX6 RecA-C domains with members of the eIF4A family and the DEAD box helicases UAP56 and DDX46. Secondary structure elements as determined from the structure are shown above the alignment. Residues that form the interface with the CNOT1 MIF4G domain are indicated by blue dots. Residues providing specificity are highlighted by dark blue dots. Residues mutated in this study are marked by red asterisks. Colors and species abbreviations are as described in Figure S2. See also Figures 5, 6, S7 and Table S1.

(J) The structure of the CNOT1-DDX6 complex described in this study together with the available structures of CNOT1 bound to deadenylases and of DDX6 in complex with decapping factors, present snapshots of consecutive steps of the 5'-to-3' mRNA decay pathway. Together with the structure of the CN9BD-CNOT9 complex, these structures establish a chain of physical interactions to describe how the CCR4-NOT

complex is recruited to miRNA targets and enrolls DDX6, which in turn represses translation and recruits the decapping machinery. See also Figure 5J. The letters in italics indicate the interactions supported by crystal or NMR structures. (a) Structure of human AGO2 bound to tryptophan (Schirle and MacRae, 2012); (b,c,d) Structures described in the present study. (e,f) Structure of the human CNOT1 MIF4G domain bound to CAF1 (Petit et al., 2012) and of *S. cerevisiae* Not1 bound to the Caf1-Ccr4 complex (Basquin et al., 2012). (g) Structure of DDX6 bound to EDC3 (Sharif et al., 2013; Tritschler et al., 2009). (h) Structure of EDC3 bound to an helical leucine rich motif (HLM) present in DCP1 in metazoans (Fromm et al., 2012). (i) Structure of the *S. pombe* Dcp1-Dcp2 complex (She et al., 2008).

Table S1. Mutants and antibodies used in this study

Name	*Hs* CNOT1 (1-2376)	*Dm* NOT1 (1-2505)	Location
CNOT1-N	1–1089	1–1148	
CNOT1-M	1085–1605	1147–1717	
CNOT1-C	1595–2376	1710–2505	
CNOT1 MIF4G	1093–1317	1152–1377	
CNOT1 SHD	1842-2353	1963–2478	
	1318–1605		
	1386–1535		
CBD/DUF3819	1356–1588	1467–1704	
	1356–1561		
	1356–1607		
	1356–1628		
	K1426S	K1537S	Interface with CNOT9
	G1451Y	G1562E, G1562Y	Interface with CNOT9
	R1458A	K1569A	Interface with CNOT9
	Q1549A	Q1658A	Interface with CNOT9
3xM	K1426S, G1451Y, R1458A	K1537S,G1562E,Q1658A	Interface with CNOT9
4xM	K1426S, G1451Y, R1458A, Q1549A		Interface with CNOT9
5xM	I1423D, K1426S, G1451Y, R1458A, Q1549A		Interface with CNOT9
Δhl	Δ1097–1110		Interface with DDX6
Δhl,N,F	Δhl,N1144A,F1145A		Interface with DDX6
Δhl,R1138A	Δhl,R1138A		Interface with DDX6
Δhl,R,N,F	Δhl,R1138A,N1144A,F1145A		Interface with DDX6

Name	*Hs* CNOT9 (1-299)	*Dm* NOT9 (1-304)	Location
ARM	19–285	25-291	
H58A	H58A	N63A	Interface with CNOT1
F60A	F60A	F65A	Interface with CNOT1
A64Y	A64Y	C69Y	Interface with CNOT1
V71Y	V71Y	V67Y	Interface with CNOT1
Quadruple Mut1	H58A,F60A,A64Y,V71Y	N63A,F65A,C69Y,V67Y	Interface with CNOT1
Mut2		N63A,F65A,C69Y	
P1a	Y203A	Y208A	Pocket 1
P1b	R244A	R249A	Pocket 1
P1c	A248F	A235F	Pocket 1
P1	Y203A,R244A	Y208A,R249A	Pocket 1
P1d	Y203A,A248F	Y208A,A253F	Pocket 1

P2a	P165G		Pocket 2
P2b	R205A		Pocket 2
P2c	H208A		Pocket 2
P2	R205A,H208A	R210A,H213A	Pocket 2
P1+2a	Y203A,R205A,H208A,R244A		Pocket 1+2
P1+2b	Y203A,R205A,H208A,A248F	Y208A,R210A,H253F,A253F	Pocket 1+2
P1+2c	P165G,Y203A,R205A, R244A		Pocket 1+2

Name	***Hs* DDX6 (1-472)** NCBI: NM_004397.3	**Location**
RecA-N	85–295	
RecA-C	296–463	
R375A	R375A	Interface with CNOT1
Q322A	Q322A	Interface with CNOT1
N324A	N324A	Interface with CNOT1
Q,N	Q322A,N324A	Interface with CNOT1

Antibody	**Source**	**Catalog Number**	**Dilution**	**Monoclonal/ Polyclonal**
Anti-HA-HRP	Roche	12 013 819 001	1:5,000	Monoclonal
Anti-GFP (for western blotting)	Roche	11 814 460 001	1:2,000	Monoclonal
Anti-GFP (for immunoprecipitations)	In house			Rabbit polyclonal
Anti-tubulin	Sigma *Aldrich*	T6199	1:5,000	Monoclonal
Anti-mouse-HRP	GE Healthcare	NA931V	1:10,000	Polyclonal
Anti-rabbit-HRP	GE Healthcare	NA934V	1:10,000	Polyclonal
Anti-DDX6	Bethyl Laboratories	A300-461A	1:3,000	Rabbit polyclonal
Anti-*Hs* CAF1	In house		1:2,000	Rabbit polyclonal
Anti-*Hs* CNOT1	In house		1:1,000	Rabbit polyclonal
Anti-*Hs* CNOT9	Proteintech	22503-1-AP	1:1,000	Rabbit polyclonal
Anti-*Hs* AGO2	Sigma *Aldrich*	SAB4200085	1:2,000	Monoclonal
Anti-*Hs* GW182 (TNRC6A)	Bethyl Laboratories	A302-329A	1:1,000	Rabbit polyclonal
Anti-V5	AbD Serotec	MCA1360GA	1:5,000	Monoclonal

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Plasmids

Plasmids expressing epitope-tagged proteins in human and *Dm* S2 cells have been previously described (Bawankar et al., 2013; Braun et al., 2011; Tritschler et al., 2009). For expression in *E. coli*, the DNAs coding for the human CNOT1 CN9BD domain (residues 1356–1607) and the ARM-repeat domain of CNOT9 (residues 19–285) were amplified by PCR using total human cDNA as the template and inserted between the XhoI and BamHI restriction sites of the pnYC-NpM (which provides an N-terminal MBP tag followed by a HRV3C protease cleavage site; Diebold et al., 2011) and pnEA-NpH (which provides an N-terminal 6xHis tag, followed by a HRV3C protease cleavage site; Diebold et al., 2011) vectors, respectively. Plasmids for expression of the CNOT1 MIF4G domain (residues 1093–1317) and the DDX6 RecA-C domain (residues 296–472) were described previously (Petit et al., 2013; Tritschler et al., 2009). DNA fragments encoding the TNRC6 silencing domains (TNRC6A residues 1462–1962 and TNRC6C residues 1261–1690) were amplified by PCR and inserted between the XhoI and AvrII restriction sites of the pnYC-NpM vector. A non-cleavable C-terminal 6x-His tag was introduced by PCR. Mutations were introduced using the QuikChange mutagenesis kit (Stratagene) with the appropriate oligonucleotides and confirmed by sequencing. The mutations used in this study are described in Table S1.

Protein Expression and Purification

The CNOT1 CN9BD domain and the CNOT9 ARM domain were co-expressed in *E. coli* BL21 Star (DE3) cells (Invitrogen) and induced at an OD_{600} of 0.6 with 1 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG). Following induction the proteins were expressed overnight at 20^{o}C. The CN9BD-CNOT9 ARM complex was purified in

lysis buffer (50 mM potassium phosphate (pH 7.5), 300 mM NaCl, 2 mM ß-mercaptoethanol) supplemented with protease inhibitors (Roche complete EDTA-free, 1 tablet per 50 ml of lysis buffer), lysozyme (SIGMA, 1 mg/ml) and DNase I (10 µg/ml), using amylose resin (New England Biolabs Inc.) to capture MBP tagged CNOT1 CN9BD as an initial purification step. The complex was digested with HRV3C protease while dialyzing into 20 mM HEPES buffer (pH 7.5), 150 mM NaCl and 1 mM DTT. The complex was further purified on a heparin column (HiTrap Heparin HP 5 ml, GE Healthcare) and a final size exclusion chromatography step (HiLoad Superdex 200 26/60 column, GE Healthcare) equilibrated in storage buffer (10 mM HEPES (pH 7.5), 150 mM NaCl and 1 mM DTT). The protein complex was concentrated to 5 mg/ml prior to use for crystallization.

Structure Solution and Refinement

Diffraction data were collected at a wavelength of 1.000 Å and processed with XDS (Kabsch, 2010). The structures were solved by molecular replacement in PHASER (McCoy et al., 2007) using the CNOT9 monomer (PDB ID code 2FV2; Garces et al., 2007) as a search model for the CN9BD-CNOT9 complex (1.65 Å resolution). For the MIF4G-DDX6 complex (1.75 Å resolution), previously determined structures of the isolated domains were used as search models (PDB ID code 4GML; Petit et al., 2012; PDB ID code 2WAX; Tritschler et al., 2009). The structures were then built automatically using the PHENIX AutoSol wizard (Terwilliger et al., 2009) and BUCCANEER (Cowtan, 2006) from the CCP4 package (Winn et al., 2011). The models were subsequently improved manually in COOT (Emsley et al., 2010) and refined using PHENIX (Afonine et al., 2012). Final refinement rounds were carried out in PHENIX, refining TLS parameters in addition to individual B-factors and including hydrogens in the riding positions. The final model of the CN9BD-CNOT9

complex was then used as a molecular replacement model for the complex at 2.05 Å resolution using PHASER. This structure contains additional free tryptophan as a ligand and was refined without the use of hydrogens. All models were subsequently improved by iterative cycles of refinement and building using PHENIX and COOT. The correct stereochemical properties of the structures were verified with MOLPROBITY (Chen et al., 2010). Figures were generated in PyMOL (http://www.pymol.org). The data collection and refinement statistics are summarized in Table 1.

***In vitro* Pulldown Assays**

The TNRC6 silencing domains were expressed overnight in *E. coli* BL21 (DE3) Star cells (Invitrogen) at 20^{o}C. All purification steps were performed on ice in lysis buffer (50 mM potassium phosphate (pH 7.5), 300 mM NaCl, 2 mM ß-mercaptoethanol), supplemented with protease inhibitors (Roche complete EDTA-free, 2 tablets per 50 ml of lysis buffer), lysozyme (1 mg/ml) and DNAse I (5 µg/ml). The bacterial lysates were incubated with amylose resin for 1 hour and eluted in lysis buffer supplemented with 30 mM maltose. The silencing domain was subsequently loaded onto a Ni-column (HiTrap IMAC HP; GE Healthcare) to separate C-terminally truncated fragments from the full-length protein. The fusion protein was dialyzed into HEPES buffer (20 mM HEPES pH (7.5), 150 mM NaCl, 1 mM DTT) and further purified by size-exclusion chromatography (HiLoad Superdex 200 26/60 column, GE Healthcare). CNOT9 mutants were purified as described for the wild-type protein.

MBP pulldown assays were performed in a final volume of 100 µl of PBS buffer using 6.25 µM of purified silencing domain and CNOT9 (wild-type and mutants). The purified proteins were incubated for 30 min at room temperature before adding 50 µl (50% slurry) of amylose resin followed by another 30 min incubation. The beads were

washed four times with PBS buffer and eluted in elution buffer (PBS supplemented with 30 mM maltose). The eluted proteins were precipitated with TCA (trichloroacetic acid) and analyzed on a 14% SDS-PAGE.

For the CNOT9-CNOT1 pulls downs shown in Figures S1B and S5C,D, His-tagged *Hs* CNOT9 (residues 19–285, wild-type or mutants) and MBP-tagged *Hs* CNOT1 (wild-type, fragments or mutants) were co-expressed in *E. coli* BL21 (DE3) star cells. The cells were lysed in suspension buffer containing 50 mM Tris (pH 7.5), 150 mM NaCl, 0.1% Triton X-100 supplemented with protease inhibitors, lysozyme, DNAse I and 10% glycerol. Cell lysates were sonicated four times for 30 s and cleared by centrifugation at 20,000 ×*g*. The cleared supernatants were incubated with 50 µl (50% slurry) of amylose resin for 45 min at 25^{o}C. The beads were washed four times with suspension buffer, bound proteins were eluted using 2x protein sample buffer and analyzed on a 14% SDS-PAGE.

The DDX6-MIF4G pulldowns were performed at room temperature with purified proteins in binding buffer containing 25 mM HEPES (pH 7.0), 20 mM NaCl and 2mM DTT. The proteins were incubated for 30 min before adding 50 µl (50% slurry) of amylose resin at room temperature, followed by another 30 min incubation. Proteins were eluted and analyzed as described above.

Complementation Assays

Plasmids expressing short-hairpin RNAs (shRNAs) for knockdowns are derived from the pSUPER plasmid containing the puromycin-resistance gene for selection. The vector backbone was a kind gift from O. Mühlemann (University of Bern). The 19 nt target sequences are as follows: control ATTCTCCGAACGTGTCACG, CNOT1 ATTCAACATTCCCTTATA, and DDX6 GCAGAAACCCTATGAGATT. HeLa cells were grown in DMEM supplemented with 10% heat inactivated fetal calf serum,

2 mM L-Glutamine, 100 U/ml penicillin and 100 mg/ml streptomycin. Cells were transfected in 6-well plates using Lipofectamine 2000 (Invitrogen) according to the manufacturers protocol. Transfection mixtures contained 4 μg of plasmids expressing the relevant shRNA. Twenty-four hours after transfection, cells were selected in medium supplemented with 1.5 μg/ml puromycin. After 1 day of selection, cells were counted and seeded in 12-well plates in medium without puromycin for recovery. Twenty-four hours after reseeding, cells were re-transfected using Lipofectamine 2000. The transfection mixtures contained 0.36 μg of plasmids expressing the relevant shRNA, 0.12 μg of the reporter pSiCHECK-8xLet7 plasmid or the corresponding reporter carrying mutations in the Let-7 binding sites and 0.12 μg of the pEGFP-N3-F-Luc transfection control. For the CNOT1 complementation assays, the transfection mixtures contained either 1 μg pT7-EGFP-CNOT1 (wild-type or mutants) or pT7-EGFP as a control. For the DDX6 complementation, the transfection mixtures contained either 0.4 μg pCIneo-λN-HA-MBP or pCIneo-λN-HA-DDX6 (wild-type or mutants). Cell were harvested 72 hours after the second transfection. R-Luc and F-Luc activities were measured using the Dual-Luciferase Reporter Assay System (Promega).

Supplemental references

Afonine, P.V., Grosse-Kunstleve, R.W., Echols, N., Headd, J.J., Moriarty, N.W., Mustyakimov, M., Terwilliger, T.C., Urzhumtsev, A., Zwart, P.H. and Adams, P.D. (2012). Towards automated crystallographic structure refinement with phenix.refine. Acta Crystallogr. D Biol. Crystallogr. *68*, 352–367.

Basquin, J., Roudko, V.V., Rode, M., Basquin, C., Séraphin, B., and Conti, E. (2012). Architecture of the nuclease module of the yeast Ccr4-not complex: the Not1-Caf1-Ccr4 interaction. Mol. Cell *48*, 207–218.

Chen, V.B., Arendall, W.B., Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr. D Biol. Crystallogr. *66*, 12–21.

Cowtan, K. (2006). The Buccaneer software for automated model building. 1. Tracing protein chains. Acta Crystallogr. D Biol. Crystallogr. *62*, 1002–1011.

Diebold, M.L., Fribourg, S., Koch, M., Metzger, T., and Romier, C. (2011). Deciphering correct strategies for multiprotein complex assembly by co-expression: application to complexes as large as the histone octamer. J. Struct. Biol. *175*, 178–188.

Emsley, P., Lohkamp, B., Scott, W.G. and Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr. D Biol. Crystallogr. *66*, 486–501.

Fromm, S.A., Truffault, V., Kamenz, J., Braun, J.E., Hoffmann, N.A., Izaurralde, E., and Sprangers, R. (2012). The structural basis of Edc3- and Scd6-mediated activation of the Dcp1:Dcp2 mRNA decapping complex. EMBO J. *31*, 279–290.

Gouet, P., Robert, X., and Courcelle E. (2003). ESPript/ENDscript: extracting and rendering sequence and 3D information from atomic structures of proteins. Nucleic Acids Res. *31*, 3320–3323.

Kabsch, W. (2010). XDS. Acta Crystallogr. D Biol. Crystallogr. *66*, 125–132.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. J. Appl. Crystallogr. *40*, 658–674.

Sharif, H., Ozgur, S., Sharma, K., Basquin, C., Urlaub, H., and Conti, E. (2013). Structural analysis of the yeast Dhh1-Pat1 complex reveals how Dhh1 engages Pat1, Edc3 and RNA in mutually exclusive interactions. Nucleic Acids Res. *41*, 8377–8390.

She, M., Decker, C.J., Svergun, D.I., Round, A., Chen, N., Muhlrad, D., Parker, R., and Song, H. (2008). Structural basis of Dcp2 recognition and activation by Dcp1. Mol. Cell *29*, 337–349.

Terwilliger, T.C., Adams, P.D., Read, R.J., Mccoy, A.J., Moriarty, N.W., Grosse-Kunstleve, R.W., Afonine, P.V., Zwart, P.H., and Hung, L.W. (2009). Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. Acta Crystallogr. D Biol. Crystallogr. *65*, 582–601.

Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G.W., McCoy, A.*, et al.* (2011). Overview of the CCP4 suite and current developments. Acta Crystallogr. D Biol. Crystallogr. *67*, 235–242.